

Vector Calculus (Maths 214)

Theodore Voronov

January 20, 2003

Contents

1	Recollection of differential calculus in $\mathbb{R}^n$	3
1.1	Points and vectors	3
1.2	Velocity vector	6
1.3	Differential of a function	9
1.4	Changes of coordinates	15
2	Line integrals and 1-forms	20
3	Algebra of forms	24
3.1	Jacobian	24
3.2	Rules of exterior multiplication	26
4	Exterior derivative	27
4.1	Dependence of line integrals on paths	27
4.2	Exterior derivative: construction	27
4.3	Main properties and examples of calculation	28
5	Stokes's theorem	29
5.1	Integration of k -forms	29
5.2	Stokes's theorem: statement and examples	34
5.3	A proof for a simple case	39
6	Classical integral theorems	41
6.1	Forms corresponding to a vector field	41
6.2	The Ostrogradski–Gauss and classical Stokes theorems	46

Introduction

Vector calculus develops on some ideas that you have learned from elementary multivariate calculus. Our main task is to develop the geometric tools. The central notion of this course is that of a *differential form* (shortly, *form*).

VISIT...

LANZAROTE
Caliente.COM

Example 1. The expressions

$$2dx + 5dy - dz$$

and

$$dx dy + e^x dy dz$$

are examples of differential forms.

In fact, the former expression above is an example of what is called a “1-form”, while the latter is an example of a “2-form”. (You can guess what 1 and 2 stand for.)

You will learn the precise definition of a form pretty soon; meanwhile I will give some more examples in order to demonstrate that to a certain extent this object is already familiar.

Example 2. In the usual integral over a segment in $\mathbb{R}$, e.g.,

$$\int_0^{2\pi} \sin x \, dx,$$

the expression $\sin x \, dx$ is a 1-form on $[0, 2\pi]$ (or on $\mathbb{R}$).

Example 3. The total differential of a function in $\mathbb{R}^3$ (if you know what it is),

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy + \frac{\partial f}{\partial z} dz,$$

is a 1-form in $\mathbb{R}^3$.

Example 4. When you integrate a function over a bounded domain in the plane:

$$\int_D f(x, y) \, dx dy$$

the expression under the integral, $f(x, y) \, dx dy$, is a 2-form in D .

We can conclude that a *form* is a linear combination of differentials or their products. Of course, we need to know the algebraic rules of handling these products. This will be discussed in due time.

When we will learn how to handle forms, this, in particular, will help us a lot with integrals.

The central statement about forms is the so-called ‘general (or generalized) Stokes theorem’. You should be familiar with what turns out to be some of its instances:

Example 5. In elementary multivariate calculus **Green’s formula** in the plane is considered:

$$\oint_C Pdx + Qdy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy,$$

where D is a domain bounded by a contour C . (The symbol $\oint$ is used for integrals over “closed contours”.)

Example 6. The **Newton–Leibniz formula** or the “fundamental theorem of calculus”:

$$F(b) - F(a) = \int_a^b F'(x) dx.$$

Here the boundary of a segment $[a, b]$ consists of two points b, a . The difference $F(b) - F(a)$ should be regarded as an “integral” over these points (taken with appropriate signs).

The generalized Stokes theorem embraces the two statements above as well as many others, which have various traditional names attached to them. It reads as follows:

Theorem.

$$\oint_{\partial M} \omega = \int_M d\omega.$$

Here ω is a differential form, M is an “oriented manifold with boundary”, $d\omega$ is the “exterior differential” of ω , ∂M is the “boundary” of M . Or, rather, we shall consider a version of this theorem with M replaced by a so-called “chain” and ∂M replaced by the “boundary” of this chain.

Our task will be to make a precise meaning of these notions.

Remark. “Vector calculus” is the name for this course, firstly, because vectors play an important role in it, and, secondly, because of a tradition. In expositions that are now obsolete, the central place was occupied by vector fields in “space” (that is, $\mathbb{R}^3$) or in the “plane” (that is, $\mathbb{R}^2$). Approach based on forms clarifies and simplifies things enormously. It allows to generalize the calculus to arbitrary $\mathbb{R}^n$ (and even further to general differentiable manifolds). The methods of the theory of differential forms nowadays are used almost everywhere in mathematics and its applications, in particular in physics and in engineering.

1 Recollection of differential calculus in $\mathbb{R}^n$

1.1 Points and vectors

Let us recall that $\mathbb{R}^n$ is the set of arrays of real numbers of length n :

$$\mathbb{R}^n = \{(x^1, x^2, \dots, x^n) \mid x^i \in \mathbb{R}, i = 1, \dots, n\}. \quad (1)$$

Here the superscript i is not a power, but simply an index. We interpret the elements of $\mathbb{R}^n$ as *points* of an “ n -dimensional space”. For points we use boldface letters (or the underscore, in hand-writing): $\mathbf{x} = (x^1, x^2, \dots, x^n)$ or $\underline{x} = (x^1, x^2, \dots, x^n)$. The numbers x^i are called the *coordinates* of the point $\mathbf{x}$. Of course, we can use letters other than $\mathbf{x}$, e.g., $\mathbf{a}$, $\mathbf{b}$ or $\mathbf{y}$, to denote points. Sometimes we also use capital letters like $A, B, C, \dots, P, Q, \dots$. A lightface letter with an index (e.g., y^i) is a generic notation for a coordinate of the corresponding point.

Example 1.1. $\mathbf{a} = (2, 5, -3) \in \mathbb{R}^3$, $\mathbf{x} = (x, y, z, t) \in \mathbb{R}^4$, $P = (1, -1) \in \mathbb{R}^2$ are points in $\mathbb{R}^3$, $\mathbb{R}^4$, $\mathbb{R}^2$, respectively. Here $a^1 = 2$, $a^2 = 5$, $a^3 = -5$; $x^1 = x$, $x^2 = y$, $x^3 = z$, $x^4 = t$; $P^1 = 1$, $P^2 = -1$. Notice that coordinates can be fixed numbers or variables.

In the examples, $\mathbb{R}^n$ often will be $\mathbb{R}^1$, $\mathbb{R}^2$ or $\mathbb{R}^3$ (maybe $\mathbb{R}^4$), but our theory is good for any n . We shall often use the “standard” coordinates x, y, z in $\mathbb{R}^3$ instead of x^1, x^2, x^3 .

Elements on $\mathbb{R}^n$ can also be interpreted as *vectors*. This you should know from linear algebra. Vectors can be added and multiplied by numbers. There is a distinguished vector “zero”: $\mathbf{0} = (0, \dots, 0)$.

Example 1.2. For $\mathbf{a} = (0, 1, 2)$ and $\mathbf{b} = (2, 3, -2)$ we have $\mathbf{a} + \mathbf{b} = (0, 1, 2) + (2, 3, -2) = (2, 4, 0)$. Also, $5\mathbf{a} = 5(0, 1, 2) = (5, 5, 10)$.

All the expected properties are satisfied (e.g., the commutative and associative laws for the addition, the distributive law for the multiplication by numbers).

Vectors are also denoted by letters with an arrow: $\overrightarrow{a} = (a^1, a^2, \dots, a^n) \in \mathbb{R}^n$. We refer to coordinates of vectors also as to their *components*.

For a time being the distinction of points and vectors is only mental.

We want to introduce two operations involving points and vectors.

Definition 1.1. For a point $\mathbf{x}$ and a vector $\mathbf{a}$ (living in the same $\mathbb{R}^n$), we define their *sum*, which is a point (by definition), as $\mathbf{x} + \mathbf{a} := (x^1 + a^1, x^2 + a^2, \dots, x^n + a^n)$. For two points $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, we define their *difference* as a vector (by definition), denoted either as $\mathbf{y} - \mathbf{x}$ or $\overrightarrow{\mathbf{x}\mathbf{y}}$, and $\mathbf{y} - \mathbf{x} = \overrightarrow{\mathbf{x}\mathbf{y}} := (y^1 - x^1, y^2 - x^2, \dots, y^n - x^n)$.

Example 1.3. Let $A = (1, 2, 3)$, $B = (-1, 0, 7)$. Then $\overrightarrow{AB} = (-2, -2, 4)$.

(From the viewpoint of arrays, the operations introduced above are no different from the addition or subtraction of vectors. The difference comes from our mental distinction of points and vectors.)

“Addition of points” or “multiplication of a point by a number” are *not* defined. Please note this.

Remark 1.1. Both points and vectors are represented by the same type of arrays in $\mathbb{R}^n$. Their distinction will become very important later.

The most important properties of the addition of a point and a vector, and of the subtraction of two points, are contained in the formulae

$$\overrightarrow{AA} = \mathbf{0}, \quad \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}; \quad (2)$$

$$\text{if } P + \mathbf{a} = Q, \text{ then } \mathbf{a} = \overrightarrow{PQ}. \quad (3)$$

They reflect our intuitive understanding of vectors as “directed segments”.

Example 1.4. Consider the point $O = (0, \dots, 0) \in \mathbb{R}^n$. For an arbitrary vector $\mathbf{r}$, the coordinates of the point $\mathbf{x} = O + \mathbf{r}$ are equal to the respective coordinates of the vector $\mathbf{r}$: $\mathbf{x} = (x^1, \dots, x^n)$ and $\mathbf{r} = (x^1, \dots, x^n)$.

The vector $\mathbf{r}$ such as in the example is called the *position vector* or the *radius-vector* of the point $\mathbf{x}$. (Or, in greater detail: $\mathbf{r}$ is the *radius-vector of $\mathbf{x}$ w.r.t. an origin O* .) Points are frequently specified by their radius-vectors. This presupposes the choice of O as the “standard origin”. (There is a temptation to identify points with their radius-vectors, which we will resist in view of the remark above.)

Let us summarize. We have considered $\mathbb{R}^n$ and interpreted its elements in two ways: as points and as vectors. Hence we may say that we dealing with the two copies of $\mathbb{R}^n$:

$$\mathbb{R}^n = \{\text{points}\}, \quad \mathbb{R}^n = \{\text{vectors}\}$$

Operations with vectors: multiplication by a number, addition. Operations with points and vectors: adding a vector to a point (giving a point), subtracting two points (giving a vector).

$\mathbb{R}^n$ treated in this way is called an *n-dimensional affine space*. (An “abstract” affine space is a pair of sets, the set of points and the set of vectors so that the operations as above are defined axiomatically.) Notice that vectors in an affine space are also known as “free vectors”. Intuitively, they are not fixed at points and “float freely” in space. Later, with the introduction of so-called curvilinear coordinates, we will see the necessity of “fixing” vectors.

From $\mathbb{R}^n$ considered as an affine space we can proceed in two opposite directions:

$$\mathbb{R}^n \text{ as a Euclidean space} \Leftarrow \mathbb{R}^n \text{ as an affine space} \Rightarrow \mathbb{R}^n \text{ as a manifold}$$

What does it mean? Going to the left means introducing some extra structure which will make the geometry richer. Going to the right means forgetting about part of the affine structure; going further in this direction will lead us to the so-called “smooth (or differentiable) manifolds”.

The theory of differential forms does not require any extra geometry. So our natural direction is to the right. The Euclidean structure, however, is useful for examples and applications. So let us say a few words about it:

Remark 1.2. *Euclidean geometry.* In $\mathbb{R}^n$ considered as an affine space we can already do a good deal of geometry. For example, we can consider lines and planes, and quadric surfaces like an ellipsoid. However, we cannot discuss such things as “lengths”, “angles” or “areas” and “volumes”. To be able to do so, we have to introduce some more definitions, making $\mathbb{R}^n$ a *Euclidean space*. Namely, we define the *length* of a vector $\mathbf{a} = (a^1, \dots, a^n)$ to be

$$|\mathbf{a}| := \sqrt{(a^1)^2 + \dots + (a^n)^2}. \quad (4)$$

After that we can also define *distances* between points as follows:

$$d(A, B) := |\overrightarrow{AB}|. \quad (5)$$

One can check that the distance so defined possesses natural properties that we expect: is it always non-negative and equals zero only for coinciding points; the distance from A to B is the same as that from B to A (symmetry); also, for three points, A , B and C , we have $d(A, B) \leq d(A, C) + d(C, B)$ (the “triangle inequality”). To define angles, we first introduce the *scalar product* of two vectors

$$(\mathbf{a}, \mathbf{b}) := a^1 b^1 + \dots + a^n b^n. \quad (6)$$

Thus $|\mathbf{a}| = \sqrt{(\mathbf{a}, \mathbf{a})}$. The scalar product is also denoted by a dot: $\mathbf{a} \cdot \mathbf{b} = (\mathbf{a}, \mathbf{b})$, and hence is often referred to as the “dot product”. Now, for nonzero vectors we define the angle between them by the equality

$$\cos \alpha := \frac{(\mathbf{a}, \mathbf{b})}{|\mathbf{a}||\mathbf{b}|}. \quad (7)$$

The angle itself is defined up to an integral multiple of 2π . For this definition to be consistent we have to ensure that the r.h.s. of (7) does not exceed 1 by the absolute value. This follows from the inequality

$$(\mathbf{a}, \mathbf{b})^2 \leq |\mathbf{a}|^2 |\mathbf{b}|^2 \quad (8)$$

known as the *Cauchy–Bunyakovsky–Schwarz inequality* (various combinations of these three names are applied in different books). One of the ways of proving (8) is to consider the scalar square of the linear combination $\mathbf{a} + t\mathbf{b}$, where $t \in \mathbb{R}$. As $(\mathbf{a} + t\mathbf{b}, \mathbf{a} + t\mathbf{b}) \geq 0$ is a quadratic polynomial in t which is never negative, its discriminant must be less or equal zero. Writing this explicitly yields (8) (check!). The triangle inequality for distances also follows from the inequality (8).

1.2 Velocity vector

The most important example of vectors for us is their occurrence as velocity vectors of parametrized curves. Consider a map $t \mapsto \mathbf{x}(t)$ from an open interval of the real line to $\mathbb{R}^n$. Such map is called a *parametrized curve* or a *path*. We will often omit the word “parametrized”.

Remark 1.3. There is another meaning of the word “curve” when it is used for a set of points line a straight line or a circle. A parametrized curve is a map, not a set of points. One can visualize it as a set of points given by its image *plus* a law according to which this set is travelled along in “time”.

Example 1.5. A straight line l in $\mathbb{R}^n$ can be specified by a point on l line and a nonzero vector in the direction of l . Hence we can make it into a parametrized curve by introducing the equation

$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}.$$

In the coordinates we have $x^i = x_0^i + tv^i$. Here t runs over $\mathbb{R}$ (infinite interval) if we want to obtain the whole line, not just a segment.

Example 1.6. A straight line in $\mathbb{R}^3$ in the direction of the vector $\mathbf{v} = (1, 0, 2)$ through the point $\mathbf{x}_0 = (1, 1, 1)$:

$$\mathbf{x}(t) = (1, 1, 1) + t(1, 0, 2)$$

or

$$\begin{aligned} x &= 1 + t \\ y &= 1 \\ z &= 1 + 2t. \end{aligned}$$

Example 1.7. The graph of the function $y = x^2$ (a parabola in $\mathbb{R}^2$) can be made a parametrized curve by introducing a parameter t as

$$\begin{aligned} x &= t \\ y &= t^2. \end{aligned}$$

Example 1.8. The following parametrized curve:

$$\begin{aligned} x &= \cos t \\ y &= \sin t, \end{aligned}$$

where $t \in \mathbb{R}$, describes a unit circle with center at the origin, which we go around infinitely many times (with constant speed) if $t \in \mathbb{R}$. If we specify some interval $(a, b) \subset [0, 2\pi]$, then we obtain just an arc of the circle.

Definition 1.2. The *velocity vector* (or, shortly, the *velocity*) of a curve $\mathbf{x}(t)$ is the vector denoted $\dot{\mathbf{x}}(t)$ or $d\mathbf{x}/dt$, where

$$\dot{\mathbf{x}}(t) = \frac{d\mathbf{x}}{dt} := \lim_{h \rightarrow 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}. \quad (9)$$

Notice that the difference $\mathbf{x}(t+h) - \mathbf{x}(t)$ is a vector, so the velocity vector is indeed a vector in $\mathbb{R}^n$. It is convenient to visualize $\dot{\mathbf{x}}(t)$ as being attached to the corresponding point $\mathbf{x}(t)$. As the directed segment $\mathbf{x}(t+h) - \mathbf{x}(t)$ lies on a secant, the velocity vector lies on the *tangent line* to our curve at the point $\mathbf{x}(t)$ (“the limit position of secants through the point $\mathbf{x}(t)$ ”). From the definition immediately follows that

$$\dot{\mathbf{x}}(t) = \left(\frac{dx^1}{dt}, \dots, \frac{dx^n}{dt} \right) \quad (10)$$

in the coordinates. (A curve is *smooth* if the velocity vector exists. In the sequel we shall use smooth curves without special explication.)

Example 1.9. For a straight line parametrized as in Example 1.5 we get $\mathbf{x}(t+h) - \mathbf{x}(t) = \mathbf{x}_0 + (t+h)\mathbf{v} - \mathbf{x}_0 - t\mathbf{v} = h\mathbf{v}$, hence $\dot{\mathbf{x}} = \mathbf{v}$ (a constant vector).

Example 1.10. In Example 1.6 we get $\dot{\mathbf{x}} = (1, 0, 2)$.

Example 1.11. In Example 1.7 we get $\dot{\mathbf{x}}(t) = (1, 2t)$. It is instructive to sketch a picture of the curve and plot the velocity vectors at $t = 0, 1, -1, 2, -2$, drawing them as attached to the corresponding points.

Example 1.12. In Example 1.8 we get $\dot{\mathbf{x}}(t) = (-\sin t, \cos t)$. Again, it is instructive to sketch a picture. (Plot the velocity vectors at $t = 0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$.)

Example 1.13. Consider the parametrized curve $x = 2 \cos t, y = 2 \sin t, z = t$ in $\mathbb{R}^3$ (representing a round helix). Then

$$\dot{\mathbf{x}} = (-2 \sin t, 2 \cos t, 1).$$

(Make a sketch!)

The velocity vector is a feature of a parametrized curve as a map, not of its image (a “physical” curve as a set of points in space). If we will change the parametrization, the velocity will change:

Example 1.14. In Example 1.8 we can introduce a new parameter s so that $t = 5s$. Hence

$$\begin{aligned} x &= \cos 5s \\ y &= \sin 5s \end{aligned}$$

will be the new equation of the curve. Then

$$\frac{d\mathbf{x}}{ds} = (-5 \sin 5s, 5 \cos 5s) = 5 \frac{d\mathbf{x}}{dt}.$$

In general, for an arbitrary curve $t \mapsto \mathbf{x}(t)$ we obtain

$$\frac{d\mathbf{x}}{ds} = \frac{dt}{ds} \frac{d\mathbf{x}}{dt} \quad (11)$$

if we introduce a new parameter s so that $t = t(s)$ is a function of s . We always assume that the change of parameter is invertible and $dt/ds \neq 0$.

Notice that the velocity is only changed by the multiplication by a nonzero scalar factor, hence its direction is not changed (only the “speed” with which we move along the curve changes). In particular, the tangent line to a curve does not depend on parametrization.

1.3 Differential of a function

Formally, the differential of a function is the following expression:

$$\boxed{df = \frac{df}{dx} dx} \quad (12)$$

for functions of one variable and

$$\boxed{df = \frac{\partial f}{\partial x^1} dx^1 + \dots + \frac{\partial f}{\partial x^n} dx^n} \quad (13)$$

for functions of many variables. Now we want to explain the meaning of the differential.

Let us start from a function $f: (a, b) \rightarrow \mathbb{R}$ defined on an interval of the real line. We shall revisit the notion of the differentiability. Fix a point x ; we want to know how the value of the function changes when we move from x to some other point $x + h$. In other words, we consider an increment $\Delta x = h$ of the independent variable and we study the corresponding increment of our function: $\Delta f = f(x + h) - f(x)$. It depends on x and on h . For “good” functions we expect that Δf is small for small Δx .

Definition 1.3. A function f is *differentiable* at x if Δf is “approximately linear” in Δx ; precisely:

$$f(x + h) - f(x) = k \cdot h + \alpha(h)h \quad (14)$$

where $\alpha(h) \rightarrow 0$ when $h \rightarrow 0$.

This can be illustrated using the graph of the function f . The coefficient k is the slope of the tangent line to the graph at the point x . The linear function of the increment h appearing in (14) is called the *differential* of f at x :

$$df(x)(h) = k \cdot h = k \cdot \Delta x. \quad (15)$$

In other words, $df(x)(h)$ is the “main (linear) part” of the increment Δf (at the point x) when $h \rightarrow 0$. Approximately $\Delta f \approx df$ when $\Delta x = h$ is small. The coefficient k is exactly the derivative of f at x . Notice that $dx = \Delta x$. Hence

$$df = k \cdot dx \quad (16)$$

where k is the derivative. (We suppressed x in the notation for df .) Thus the common notation df/dx for the derivative can be understood directly as the ratio of the differentials.

This definition of differentiability for functions of a single variable is equivalent to the one where the derivative comes first and the differential is defined later. It is worth teaching yourself to think in terms of differentials.

Example 1.15. Differentials of elementary functions:

$$\begin{aligned} d(x^n) &= nx^{n-1} dx \\ d(e^x) &= e^x dx \\ d(\ln x) &= \frac{dx}{x} \\ d(\sin x) &= \cos x dx, \end{aligned}$$

etc.

The same approach works for functions of many variables. Consider $f: U \rightarrow \mathbb{R}$ where $U \subset \mathbb{R}^n$. Fix a point $\mathbf{x} \in U$. The main difference from functions of a single variable is that the increment of $\mathbf{x}$ is now a vector: $\mathbf{h} = (h^1, \dots, h^n)$. Consider $\Delta f = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})$ for various $\mathbf{h} \in \mathbb{R}^n$. For this to make sense at least for small $\mathbf{h}$ we need the domain U where f is defined to be *open*, i.e. containing a small ball around $\mathbf{x}$ (for every $\mathbf{x} \in U$).

Definition 1.4. A function $f: U \rightarrow \mathbb{R}$ is *differentiable* at $\mathbf{x}$ if

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = A(\mathbf{h}) + \alpha(\mathbf{h})|\mathbf{h}| \quad (17)$$

where $A(\mathbf{h}) = A_1 h^1 + \dots + A_n h^n$ is a linear function of $\mathbf{h}$ and $\alpha(\mathbf{h}) \rightarrow 0$ when $\mathbf{h} \rightarrow 0$. (The function A , of course, depends on $\mathbf{x}$.) The linear function $A(\mathbf{h})$ is called the *differential* of f at $\mathbf{x}$. Notation: df or $df(\mathbf{x})$, so $df(\mathbf{x})(\mathbf{h}) = A(\mathbf{h})$. The value of df on a vector $\mathbf{h}$ is also called the *derivative of f along $\mathbf{h}$* and denoted $\partial_{\mathbf{h}} f(\mathbf{x}) = df(\mathbf{x})(\mathbf{h})$.

Example 1.16. Let $f(\mathbf{x}) = (x^1)^2 + (x^2)^2$ in $\mathbb{R}^2$. Choose $\mathbf{x} = (1, 2)$. Then $df(\mathbf{x})(\mathbf{h}) = 2h^1 + 4h^2$ (check!).

Example 1.17. Consider $\mathbf{h} = \mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0)$ (the i -th standard basis vector in $\mathbb{R}^n$). The derivative $\partial_{\mathbf{e}_i} f = df(\mathbf{x})(\mathbf{e}_i) = A_i$ is called the *partial derivative* w.r.t. x^i . The standard notation:

$$df(\mathbf{x})(\mathbf{e}_i) =: \frac{\partial f}{\partial x^i}(\mathbf{x}). \quad (18)$$

From Definition 1.4 immediately follows that partial derivatives are just the usual derivatives of a function of a single variable if we allow only one coordinate x^i to change and fix all other coordinates.

Example 1.18. Consider the function $f(\mathbf{x}) = x^i$ (the i -th coordinate). The linear function dx^i (the differential of x^i) applied to an arbitrary vector $\mathbf{h}$ is simply h^i .

From these examples follows that we can rewrite df as

$$df = \frac{\partial f}{\partial x^1} dx^1 + \dots + \frac{\partial f}{\partial x^n} dx^n, \quad (19)$$

which is the standard form. Once again: the partial derivatives in (19) are just the coefficients (depending on $\mathbf{x}$); $dx^1, dx^2, \dots$ are linear functions giving on an arbitrary vector $\mathbf{h}$ its coordinates $h^1, h^2, \dots$, respectively. Hence

$$df(\mathbf{x})(\mathbf{h}) = \partial_{\mathbf{h}} f(\mathbf{x}) = \frac{\partial f}{\partial x^1} h^1 + \dots + \frac{\partial f}{\partial x^n} h^n. \quad (20)$$

Theorem 1.1. Suppose we have a parametrized curve $t \mapsto \mathbf{x}(t)$ passing through $\mathbf{x}_0 \in \mathbb{R}^n$ at $t = t_0$ and with the velocity vector $\dot{\mathbf{x}}(t_0) = \mathbf{v}$. Then

$$\frac{d f(\mathbf{x}(t))}{dt}(t_0) = \partial_{\mathbf{v}} f(\mathbf{x}_0) = df(\mathbf{x}_0)(\mathbf{v}). \quad (21)$$

Proof. Indeed, consider a small increment of the parameter t : $t_0 \mapsto t_0 + \Delta t$. We want to find the increment of $f(\mathbf{x}(t))$. We have

$$\mathbf{x}_0 \mapsto \mathbf{x}(t_0 + \Delta t) = \mathbf{x}_0 + \mathbf{v} \cdot \Delta t + \alpha(\Delta t)\Delta t$$

where $\Delta t \rightarrow 0$. On the other hand, we have

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = df(\mathbf{x}_0)(\mathbf{h}) + \beta(\mathbf{h})|\mathbf{h}|$$

for an arbitrary vector $\mathbf{h}$, where $\beta(\mathbf{h}) \rightarrow 0$ when $\mathbf{h} \rightarrow 0$. Combining it together, for the increment of $f(\mathbf{x}(t))$ we obtain

$$\begin{aligned} f(\mathbf{x}(t_0 + \Delta t)) - f(\mathbf{x}_0) &= \\ df(\mathbf{x}_0)(\mathbf{v} \cdot \Delta t + \alpha(\Delta t)\Delta t) + \beta(\mathbf{v} \cdot \Delta t + \alpha(\Delta t)\Delta t) \cdot |\mathbf{v}\Delta t + \alpha(\Delta t)\Delta t| &= \\ df(\mathbf{x}_0)(\mathbf{v}) \cdot \Delta t + \gamma(\Delta t)\Delta t \end{aligned}$$

for a certain $\gamma(\Delta t)$ such that $\gamma(\Delta t) \rightarrow 0$ when $\Delta t \rightarrow 0$ (we used the linearity of $df(\mathbf{x}_0)$). By the definition, this means that the derivative of $f(\mathbf{x}(t))$ at $t = t_0$ is exactly $df(\mathbf{x}_0)(\mathbf{v})$. $\square$

The statement of the theorem can be expressed by a simple formula:

$$\frac{df(\mathbf{x}(t))}{dt} = \frac{\partial f}{\partial x^1} \dot{x}^1 + \dots + \frac{\partial f}{\partial x^n} \dot{x}^n. \quad (22)$$

Theorem 1.1 gives another approach to differentials: to calculate the value of df at a point $\mathbf{x}_0$ on a given vector $\mathbf{v}$ one can take an arbitrary curve passing through $\mathbf{x}_0$ at t_0 with $\mathbf{v}$ as the velocity vector at t_0 and calculate the usual derivative of $f(\mathbf{x}(t))$ at $t = t_0$.

Theorem 1.2. *For functions $f, g: U \rightarrow \mathbb{R}$, where $U \subset \mathbb{R}^n$,*

$$d(f + g) = df + dg \quad (23)$$

$$d(fg) = df \cdot g + f \cdot dg \quad (24)$$

Proof. We can prove this either directly from Definition 1.4 or using formula (21). Consider an arbitrary point $\mathbf{x}_0$ and an arbitrary vector $\mathbf{v}$ stretching from it. Let a curve $\mathbf{x}(t)$ be such that $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\dot{\mathbf{x}}(t_0) = \mathbf{v}$. Hence $d(f + g)(\mathbf{x}_0)(\mathbf{v}) = \frac{d}{dt}(f(\mathbf{x}(t)) + g(\mathbf{x}(t)))$ at $t = t_0$ and $d(fg)(\mathbf{x}_0)(\mathbf{v}) = \frac{d}{dt}(f(\mathbf{x}(t))g(\mathbf{x}(t)))$ at $t = t_0$. Formulae (23) and (24) then immediately follow from the corresponding formulae for the usual derivative (check!). $\square$

Now, almost without change the theory generalizes to functions taking values in $\mathbb{R}^m$ instead of $\mathbb{R}$. The only difference is that now the differential of a map $\mathbf{F}: U \rightarrow \mathbb{R}^m$ at a point $\mathbf{x}$ will be a linear function taking vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$ (instead of $\mathbb{R}$). For an arbitrary vector $\mathbf{h} \in \mathbb{R}^n$,

$$\mathbf{F}(\mathbf{x} + \mathbf{h}) = \mathbf{F}(\mathbf{x}) + d\mathbf{F}(\mathbf{x})(\mathbf{h}) + \beta(\mathbf{h})|\mathbf{h}| \quad (25)$$

where $\beta(\mathbf{h}) \rightarrow 0$ when $\mathbf{h} \rightarrow 0$. We have $d\mathbf{F} = (dF^1, \dots, dF^m)$ and

$$d\mathbf{F} = \frac{\partial \mathbf{F}}{\partial x^1} dx^1 + \dots + \frac{\partial \mathbf{F}}{\partial x^n} dx^n = \begin{pmatrix} \frac{\partial F^1}{\partial x^1} & \dots & \frac{\partial F^1}{\partial x^n} \\ \dots & \dots & \dots \\ \frac{\partial F^m}{\partial x^1} & \dots & \frac{\partial F^m}{\partial x^n} \end{pmatrix} \begin{pmatrix} dx^1 \\ \dots \\ dx^n \end{pmatrix}. \quad (26)$$

In this matrix notation we have to write vectors as vector-columns.

Theorem 1.1 generalizes as follows.

Theorem 1.3. *For an arbitrary parametrized curve $\mathbf{x}(t)$ in $\mathbb{R}^n$, the differential of a map $\mathbf{F}: U \rightarrow \mathbb{R}^m$ (where $U \subset \mathbb{R}^n$) maps the velocity vector $\dot{\mathbf{x}}(t)$ to the velocity vector of the curve $\mathbf{F}(\mathbf{x}(t))$ in $\mathbb{R}^m$:*

$$\frac{d\mathbf{F}(\mathbf{x}(t))}{dt} = d\mathbf{F}(\mathbf{x}(t))(\dot{\mathbf{x}}(t)). \quad (27)$$

Proof. By the definition of the velocity vector,

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \dot{\mathbf{x}}(t) \cdot \Delta t + \boldsymbol{\alpha}(\Delta t)\Delta t \quad (28)$$

where $\boldsymbol{\alpha}(\Delta t) \rightarrow 0$ when $\Delta t \rightarrow 0$. By the definition of the differential,

$$\mathbf{F}(\mathbf{x} + \mathbf{h}) = \mathbf{F}(\mathbf{x}) + d\mathbf{F}(\mathbf{x})(\mathbf{h}) + \boldsymbol{\beta}(\mathbf{h})|\mathbf{h}| \quad (29)$$

where $\boldsymbol{\beta}(\mathbf{h}) \rightarrow 0$ when $\mathbf{h} \rightarrow 0$. Plugging (28) into (29), we obtain

$$\begin{aligned} \mathbf{F}(\mathbf{x}(t+\Delta t)) &= \mathbf{F}\left(\mathbf{x} + \underbrace{\dot{\mathbf{x}}(t) \cdot \Delta t + \boldsymbol{\alpha}(\Delta t)\Delta t}_{\mathbf{h}}\right) = \mathbf{F}(\mathbf{x}) + d\mathbf{F}(\mathbf{x})(\dot{\mathbf{x}}(t)\Delta t + \boldsymbol{\alpha}(\Delta t)\Delta t) + \\ &\quad \boldsymbol{\beta}(\dot{\mathbf{x}}(t)\Delta t + \boldsymbol{\alpha}(\Delta t)\Delta t) \cdot |\dot{\mathbf{x}}(t)\Delta t + \boldsymbol{\alpha}(\Delta t)\Delta t| = \mathbf{F}(\mathbf{x}) + d\mathbf{F}(\mathbf{x})\dot{\mathbf{x}}(t) \cdot \Delta t + \boldsymbol{\gamma}(\Delta t)\Delta t \end{aligned}$$

for some $\boldsymbol{\gamma}(\Delta t) \rightarrow 0$ when $\Delta t \rightarrow 0$. This precisely means that $d\mathbf{F}(\mathbf{x})\dot{\mathbf{x}}(t)$ is the velocity vector of $\mathbf{F}(\mathbf{x})$. $\square$

As every vector attached to a point can be viewed as the velocity vector of some curve passing through this point, this theorem gives a clear geometric picture of $d\mathbf{F}$ as a linear map on vectors.

Theorem 1.4 (Chain rule for differentials). *Suppose we have two maps $\mathbf{F}: U \rightarrow V$ and $\mathbf{G}: V \rightarrow W$, where $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^m$, $W \subset \mathbb{R}^p$ (open domains). Let $\mathbf{F}: \mathbf{x} \mapsto \mathbf{y} = \mathbf{F}(\mathbf{x})$. Then the differential of the composite map $\mathbf{G} \circ \mathbf{F}: U \rightarrow W$ is the composition of the differentials of $\mathbf{F}$ and $\mathbf{G}$:*

$$d(\mathbf{G} \circ \mathbf{F})(\mathbf{x}) = d\mathbf{G}(\mathbf{y}) \circ d\mathbf{F}(\mathbf{x}). \quad (30)$$

Proof. We can use the description of the differential given by Theorem 1.3. Consider a curve $\mathbf{x}(t)$ in $\mathbb{R}^n$ with the velocity vector $\dot{\mathbf{x}}$. Basically, we need to know to which vector in $\mathbb{R}^p$ it is taken by $d(\mathbf{G} \circ \mathbf{F})$. By Theorem 1.3, it is the velocity vector to the curve $(\mathbf{G} \circ \mathbf{F})(\mathbf{x}(t)) = \mathbf{G}(\mathbf{F}(\mathbf{x}(t)))$. By the same theorem, it equals the image under $d\mathbf{G}$ of the velocity vector to the curve $\mathbf{F}(\mathbf{x}(t))$ in $\mathbb{R}^m$. Applying the theorem once again, we see that the velocity vector to the curve $\mathbf{F}(\mathbf{x}(t))$ is the image under $d\mathbf{F}$ of the vector $\dot{\mathbf{x}}(t)$. Hence $d(\mathbf{G} \circ \mathbf{F})(\dot{\mathbf{x}}) = d\mathbf{G}(d\mathbf{F}(\dot{\mathbf{x}}))$ for an arbitrary vector $\dot{\mathbf{x}}$ (we suppressed points from the notation), which is exactly the desired formula (30). $\square$

Corollary 1.1. *If we denote coordinates in $\mathbb{R}^n$ by $(x^1, \dots, x^n)$ and in $\mathbb{R}^m$ by $(y^1, \dots, y^m)$, and write*

$$d\mathbf{F} = \frac{\partial \mathbf{F}}{\partial x^1} dx^1 + \dots + \frac{\partial \mathbf{F}}{\partial x^n} dx^n \quad (31)$$

$$d\mathbf{G} = \frac{\partial \mathbf{G}}{\partial y^1} dy^1 + \dots + \frac{\partial \mathbf{G}}{\partial y^m} dy^m, \quad (32)$$

then the chain rule can be expressed as follows:

$$d(\mathbf{G} \circ \mathbf{F}) = \frac{\partial \mathbf{G}}{\partial y^1} dF^1 + \dots + \frac{\partial \mathbf{G}}{\partial y^m} dF^m, \quad (33)$$

where dF^i are taken from (31). In other words, to get $d(\mathbf{G} \circ \mathbf{F})$ we have to substitute into (32) the expression for $dy^i = dF^i$ from (31).

This can also be expressed by the following matrix formula:

$$d(\mathbf{G} \circ \mathbf{F}) = \begin{pmatrix} \frac{\partial G^1}{\partial y^1} & \dots & \frac{\partial G^1}{\partial y^m} \\ \dots & \dots & \dots \\ \frac{\partial G^p}{\partial y^1} & \dots & \frac{\partial G^p}{\partial y^m} \end{pmatrix} \begin{pmatrix} \frac{\partial F^1}{\partial x^1} & \dots & \frac{\partial F^1}{\partial x^n} \\ \dots & \dots & \dots \\ \frac{\partial F^m}{\partial x^1} & \dots & \frac{\partial F^m}{\partial x^n} \end{pmatrix} \begin{pmatrix} dx^1 \\ \dots \\ dx^n \end{pmatrix}, \quad (34)$$

i.e. if $d\mathbf{G}$ and $d\mathbf{F}$ are expressed by matrices of partial derivatives, then $d(\mathbf{G} \circ \mathbf{F})$ is expressed by the product of these matrices. This is often written as

$$\begin{pmatrix} \frac{\partial z^1}{\partial x^1} & \dots & \frac{\partial z^1}{\partial x^n} \\ \dots & \dots & \dots \\ \frac{\partial z^p}{\partial x^1} & \dots & \frac{\partial z^p}{\partial x^n} \end{pmatrix} = \begin{pmatrix} \frac{\partial z^1}{\partial y^1} & \dots & \frac{\partial z^1}{\partial y^m} \\ \dots & \dots & \dots \\ \frac{\partial z^p}{\partial y^1} & \dots & \frac{\partial z^p}{\partial y^m} \end{pmatrix} \begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \dots & \frac{\partial y^1}{\partial x^n} \\ \dots & \dots & \dots \\ \frac{\partial y^m}{\partial x^1} & \dots & \frac{\partial y^m}{\partial x^n} \end{pmatrix}, \quad (35)$$

or

$$\frac{\partial z^\mu}{\partial x^a} = \sum_{i=1}^m \frac{\partial z^\mu}{\partial y^i} \frac{\partial y^i}{\partial x^a}, \quad (36)$$

where it is assumed that the dependence of $\mathbf{y} \in \mathbb{R}^m$ on $\mathbf{x} \in \mathbb{R}^n$ is given by the map $\mathbf{F}$, the dependence of $\mathbf{z} \in \mathbb{R}^p$ on $\mathbf{y} \in \mathbb{R}^m$ is given by the map $\mathbf{G}$, and the dependence of $\mathbf{z} \in \mathbb{R}^p$ on $\mathbf{x} \in \mathbb{R}^n$ is given by the composition $\mathbf{G} \circ \mathbf{F}$.

Experience shows that it is much more easier to work in terms of differentials than in terms of partial derivatives.

Example 1.19. $d \cos^2 x = 2 \cos x d \cos x = -2 \cos x \sin x dx = -\sin 2x dx$.

Example 1.20. $d e^{x^2} = e^{x^2} d(x^2) = 2x e^{x^2} dx$.

Example 1.21. $d(x^2 + y^2 + z^2) = d(x^2) + d(y^2) + d(z^2) = 2x dx + 2y dy + 2z dz = 2(x dx + y dy + z dz)$.

Example 1.22. $d(t, t^2, t^3) = (dt, d(t^2), d(t^3)) = (dt, 2t dt, 3t^2 dt) = (1, 2t, 3t^2) dt$.

Example 1.23. $d(x + y, x - y) = (d(x + y), d(x - y)) = (dx + dy, dx - dy) = (dx, dx) + (dy, -dy) = (1, 1) dx + (1, -1) dy$.

Remark 1.4. The definition of the differential involves objects like $\alpha(\mathbf{h})|\mathbf{h}|$ and the notion of the limit in $\mathbb{R}^n$. At the first glance it may seem that the theory essentially relies on a Euclidean structure in $\mathbb{R}^n$ (the concept of “length”, as defined in Remark 1.2). However, it is not so. One can check that the notions of the limit as well as the conditions like “a function of the form $\alpha(\mathbf{h})|\mathbf{h}|$ where $\alpha(\mathbf{h}) \rightarrow 0$ when $\mathbf{h} \rightarrow 0$ ” are equivalent for all reasonable definitions of “length” in $\mathbb{R}^n$, hence are intrinsic for $\mathbb{R}^n$ and do not depend on any extra structure.

1.4 Changes of coordinates

Consider points $\mathbf{x} = (x^1, \dots, x^n) \in \mathbb{R}^n$. From now on we will call the numbers x^i the “standard coordinates” of the point $\mathbf{x}$. The reason is that we can use other coordinates to specify points. Before we give a general definition, let us consider some examples.

Example 1.24. “New coordinates” x', y' are introduced in $\mathbb{R}^2$ by the formulae

$$\begin{cases} x' = 2x - y \\ y' = x + y. \end{cases}$$

For example, if a point $\mathbf{x}$ has the standard coordinates $x = 1, y = 2$, then the new coordinates of the same point will be $x' = 0, y' = 3$. We can, conversely, express x, y in terms of x', y' :

$$\begin{cases} x = \frac{1}{3}(x' + y') \\ y = \frac{1}{3}(-x' + 2y'). \end{cases}$$

The geometric meaning of such change of coordinates is very simple. Recall that the standard coordinates of a point $\mathbf{x}$ coincide with the components of the radius-vector $\mathbf{r} = \overrightarrow{O\mathbf{x}}$: $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2$, where $\mathbf{e}_1 = (1, 0)$, $\mathbf{e}_2 = (0, 1)$ is the standard basis. “New” coordinates as above correspond to a different choice of a basis: $\mathbf{r} = x'\mathbf{e}'_1 + y'\mathbf{e}'_2$.

Problem 1.1. Find $\mathbf{e}'_1$ and $\mathbf{e}'_2$ for the example above from the condition that $x\mathbf{e}_1 + y\mathbf{e}_2 = x'\mathbf{e}'_1 + y'\mathbf{e}'_2$ for all points (x, y) and x', y' are related by the formulae above).

Another way to get new coordinates in $\mathbb{R}^n$ is to choose a “new origin” instead of $O = (0, \dots, 0)$.

Example 1.25. Define x', y' in $\mathbb{R}^2$ by the formulae

$$\begin{cases} x' = x + 5 \\ y' = y - 3. \end{cases}$$

Then the “old origin” O will have the new coordinates $x' = 5, y' = -3$. Conversely, we can find the “new origin” O' , i.e., the point that has the new coordinates $x' = 0, y' = 0$: its old coordinates are $x = -5, y = 3$.

Now we will consider a non-linear change of coordinates.

Example 1.26. Consider in $\mathbb{R}^2$ the *polar coordinates* (r, φ) so that

$$\begin{cases} x = r \cos \varphi \\ y = r \sin \varphi. \end{cases} \quad (37)$$

Let $r \geq 0$. Then

$$r = \sqrt{x^2 + y^2}. \quad (38)$$

As for the angle φ , for $(x, y) \neq (0, 0)$ it can be expressed up to integral multiples of 2π via the inverse trigonometric functions. Note that φ is not defined at the origin. The correspondence between (x, y) and (r, φ) cannot be made one-to-one in the whole plane. To define φ uniquely and not just up to 2π , we have to cut out a ray starting from the origin (then we can count angles from that ray). For example, if we cut out the positive ray of the x -axis, then we can choose $\varphi \in (0, 2\pi)$ and express φ as

$$\varphi = \arccos \frac{x}{\sqrt{x^2 + y^2}}. \quad (39)$$

Hence formulae (37) and (38,39) give mutually inverse maps $\mathbf{F}: V \rightarrow U$ and $\mathbf{G}: U \rightarrow V$, where $U = \mathbb{R}^2 \setminus \{(x, 0) \mid x \geq 0\}$ is a domain of the (x, y) -plane and $V = \{(r, \varphi) \mid r > 0, \varphi \in (0, 2\pi)\}$ is a domain of the (r, φ) -plane. Notice that we can differentiate both maps infinitely many times.

We shall call a map *smooth* if it is infinitely differentiable, i.e., if there exist partial derivatives of all orders.

Modelling on the examples above, we are going to give a general definition of a “coordinate system” (or “system of coordinates”). Notice that the example of polar coordinates shows that a “system of coordinates” should be defined for a domain of $\mathbb{R}^n$, not for the whole $\mathbb{R}^n$ (in general).

Definition 1.5. Consider an open domain $U \subset \mathbb{R}^n$. Consider also another copy of $\mathbb{R}^n$, denoted for distinction $\mathbb{R}_y^n$, with the standard coordinates $(y^1, \dots, y^n)$. A *system of coordinates* in the open domain U is given by a map $\mathbf{F}: V \rightarrow U$, where $V \subset \mathbb{R}_y^n$ is an open domain of $\mathbb{R}_y^n$, such that the following three conditions are satisfied:

- (1) $\mathbf{F}$ is smooth;
- (2) $\mathbf{F}$ is invertible;
- (3) $\mathbf{F}^{-1}: U \rightarrow V$ is also smooth.

The *coordinates* of a point $\mathbf{x} \in U$ in this system are the standard coordinates of $\mathbf{F}^{-1}(\mathbf{x}) \in \mathbb{R}_y^n$.

In other words,

$$\mathbf{F}: (y^1, \dots, y^n) \mapsto \mathbf{x} = \mathbf{x}(y^1, \dots, y^n). \quad (40)$$

Here the variables $(y^1, \dots, y^n)$ are the “new” coordinates of the point $\mathbf{x}$.

The standard coordinates in $\mathbb{R}^n$ are a particular case when the map $\mathbf{F}$ is identical. In Examples 1.24 and 1.25 we have maps $\mathbb{R}_{(x', y')}^2 \rightarrow \mathbb{R}^2$; in Example 1.26 we have a map $\mathbb{R}_{(r, \varphi)}^2 \supset V \rightarrow U \subset \mathbb{R}^2$.

Remark 1.5. Coordinate systems as introduced in Definition 1.5 are often called “curvilinear”. They, of course, also embrace “linear” changes of coordinates like those in Examples 1.24 and 1.25.

Remark 1.6. There are plenty of examples of particular coordinate systems. The reason why they are introduced is that a “good choice” of coordinates can substantially simplify a problem. For example, polar coordinates in the plane are very useful in planar problems with a rotational symmetry.

What happens with vectors when we pass to general coordinate systems in domains of $\mathbb{R}^n$?

Clearly, operations with points like taking difference, $\overrightarrow{AB} = B - A$, cannot survive nonlinear maps involved in the definition of curvilinear coordinates. The hint on what to do is the notion of the velocity vector. The slogan is: “Every vector is a velocity vector for some curve.” Hence we have to figure out how to handle velocity vectors. For this we return first to the example of polar coordinates.

Example 1.27. Consider a curve in $\mathbb{R}^2$ specified in polar coordinates as

$$\mathbf{x}(t): r = r(t), \varphi = \varphi(t). \quad (41)$$

How to find the velocity $\dot{\mathbf{x}}$? We can simply use the chain rule. The map $t \mapsto \mathbf{x}(t)$ can be considered as the composition of the maps $t \mapsto (r(t), \varphi(t))$, $(r, \varphi) \mapsto \mathbf{x}(r, \varphi)$. Then, by the chain rule, we have

$$\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} = \frac{\partial \mathbf{x}}{\partial r} \frac{dr}{dt} + \frac{\partial \mathbf{x}}{\partial \varphi} \frac{d\varphi}{dt} = \frac{\partial \mathbf{x}}{\partial r} \dot{r} + \frac{\partial \mathbf{x}}{\partial \varphi} \dot{\varphi}. \quad (42)$$

Here $\dot{r}$ and $\dot{\varphi}$ are scalar coefficients depending on t , whence the partial derivatives $\partial \mathbf{x} / \partial r$, $\partial \mathbf{x} / \partial \varphi$ are vectors depending on point in $\mathbb{R}^2$. We can compare this with the formula in the “standard” coordinates: $\dot{\mathbf{x}} = \mathbf{e}_1 \dot{x} + \mathbf{e}_2 \dot{y}$. Consider the vectors $\partial \mathbf{x} / \partial r$, $\partial \mathbf{x} / \partial \varphi$. Explicitly we have

$$\frac{\partial \mathbf{x}}{\partial r} = (\cos \varphi, \sin \varphi) \quad (43)$$

$$\frac{\partial \mathbf{x}}{\partial \varphi} = (-r \sin \varphi, r \cos \varphi), \quad (44)$$

from where it follows that these vectors make a basis at all points except for the origin (where $r = 0$). It is instructive to sketch a picture, drawing vectors corresponding to a point as starting from that point. Notice that $\partial \mathbf{x} / \partial r$, $\partial \mathbf{x} / \partial \varphi$ are, respectively, the velocity vectors for the curves $r \mapsto \mathbf{x}(r, \varphi)$ ($\varphi = \varphi_0$ fixed) and $\varphi \mapsto \mathbf{x}(r, \varphi)$ ($r = r_0$ fixed). We can conclude that for an arbitrary curve given in polar coordinates the velocity vector will have components $(\dot{r}, \dot{\varphi})$ if as a basis we take $\mathbf{e}_r := \partial \mathbf{x} / \partial r$, $\mathbf{e}_\varphi := \partial \mathbf{x} / \partial \varphi$:

$$\dot{\mathbf{x}} = \mathbf{e}_r \dot{r} + \mathbf{e}_\varphi \dot{\varphi}. \quad (45)$$

A characteristic feature of the basis $\mathbf{e}_r, \mathbf{e}_\varphi$ is that it is not “constant” but depends on point. Vectors “stuck to points” when we consider curvilinear coordinates.

Example 1.28. Question: given a vector $\mathbf{v} = 2\mathbf{e}_r + 3\mathbf{e}_\varphi$, express it in standard coordinates. The question is ill-defined unless we specify a point in $\mathbb{R}^2$ at which we consider our vector. Indeed, if we simply plug $\mathbf{e}_r, \mathbf{e}_\varphi$ from (43), (44), we get $\mathbf{v} = 2(\cos \varphi, \sin \varphi) + 3(-r \sin \varphi, r \cos \varphi)$, something with variable coefficients. Until we specify r, φ , i.e., specify a point in the plane, we cannot get numbers!

Problem 1.2. Check that solving (43), (44) for the standard basis $\mathbf{e}_1, \mathbf{e}_2$ we obtain

$$\mathbf{e}_1 = \frac{1}{r} (\mathbf{e}_r r \cos \varphi - \mathbf{e}_\varphi \sin \varphi) \quad (46)$$

$$\mathbf{e}_2 = \frac{1}{r} (\mathbf{e}_r r \sin \varphi + \mathbf{e}_\varphi \cos \varphi). \quad (47)$$

After considering these examples, we can say how vectors should be handled using arbitrary (curvilinear) coordinates. Suppose we are given a system of coordinates in a domain $U \subset \mathbb{R}^n$. We shall denote the coordinates in this system simply by $x^1, \dots, x^n$. (So x^i no longer stand for the “standard” coordinates!) Then:

- (1) there appears a “variable basis” associated with this coordinate system, which we denote $\mathbf{e}_i = \frac{\partial \mathbf{x}}{\partial x^i}$;
- (2) vectors are attached to points; every vector is specified by components w.r.t. the basis $\mathbf{e}_i$;
- (3) if we change coordinates from x^i to $x^{i'}$, then the basis transforms according (formally) to the chain rule:

$$\mathbf{e}_i = \sum \mathbf{e}_{i'} \frac{\partial x^{i'}}{\partial x^i}, \quad (48)$$

with coefficients depending on point;

- (4) the components of vectors at each point transform accordingly.

It exactly the transformation law with variable coefficients that make us consider the basis $\mathbf{e}_i$ associated with a coordinate system as “variable” and attach vectors to points.

This new approach to vectors is compatible with our original approach when we treated vectors and points simply as arrays and vectors were not attached to points.

Example 1.29. Suppose x^i are the standard coordinates in $\mathbb{R}^n$ so that $\mathbf{x} = (x^1, \dots, x^n)$. Then we can understand $\mathbf{e}_i = \partial \mathbf{x} / \partial x^i$ straightforwardly and by differentiation get at each place either 1 or 0 depending on whether we differentiate $\partial x^j / \partial x^i$ for $j = i$ or not: hence $\mathbf{e}_1 = (1, 0, \dots, 0)$, $\mathbf{e}_2 = (0, 1, 0, \dots, 0)$, $\dots$, $\mathbf{e}_n = (0, \dots, 0, 1)$. From the general rule we have recovered the standard basis in $\mathbb{R}^n$!

Remark 1.7. The “affine structure” in $\mathbb{R}^n$, i.e., the operations with points and vectors described in Section 1.1 and in particular the possibility to consider vectors independently of points (“free” vectors) is preserved under a special class of changes of coordinates, namely, those similar to Examples 1.24 and 1.25 (linear transformations and parallel translations).

With this new understanding of vectors, we can define the *velocity vector* for a parametrized curve specified w.r.t. arbitrary coordinates as $x^i = x^i(t)$ as the vector $\dot{\mathbf{x}} := \mathbf{e}_i \dot{x}^i$.

Proposition 1.1. *The velocity vector has the same appearance in all coordinate systems.*

Proof. Follows directly from the chain rule and the transformation law for the basis $\mathbf{e}_i$. $\square$

In particular, the elements of the basis $\mathbf{e}_i = \partial \mathbf{x} / \partial x^i$ (originally, a formal notation) can be understood directly as the velocity vectors of the coordinate lines $x^i \mapsto \mathbf{x}(x^1, \dots, x^n)$ (all coordinates but x^i are fixed).

Now, what happens with the differentials of maps?

Since we now know how to handle velocities in arbitrary coordinates, the best way to treat the differential of a map $\mathbf{F}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is by its action on the velocity vectors. By definition, we set

$$d\mathbf{F}(\mathbf{x}_0): \frac{d\mathbf{x}(t)}{dt}(t_0) \mapsto \frac{d\mathbf{F}(\mathbf{x}(t))}{dt}(t_0). \quad (49)$$

Now $d\mathbf{F}(\mathbf{x}_0)$ is a linear map that takes vectors attached to a point $\mathbf{x}_0 \in \mathbb{R}^n$ to vectors attached to the point $\mathbf{F}(\mathbf{x}_0) \in \mathbb{R}^m$. Using Theorem 1.3 backwards, we obtain

$$d\mathbf{F} = \frac{\partial \mathbf{F}}{\partial x^1} dx^1 + \dots + \frac{\partial \mathbf{F}}{\partial x^n} dx^n = (\mathbf{e}_1, \dots, \mathbf{e}_n) \begin{pmatrix} \frac{\partial F^1}{\partial x^1} & \dots & \frac{\partial F^1}{\partial x^n} \\ \dots & \dots & \dots \\ \frac{\partial F^m}{\partial x^1} & \dots & \frac{\partial F^m}{\partial x^n} \end{pmatrix} \begin{pmatrix} dx^1 \\ \dots \\ dx^n \end{pmatrix}, \quad (50)$$

as before, — but now these formulae are valid in all coordinate systems.

In particular, for the differential of a function we always have

$$df = \frac{\partial f}{\partial x^1} dx^1 + \dots + \frac{\partial f}{\partial x^n} dx^n, \quad (51)$$

where x^i are arbitrary coordinates. The form of the differential does not change when we perform a change of coordinates.

Example 1.30. Consider the following function in $\mathbb{R}^2$: $f = r^2 = x^2 + y^2$. We want to calculate its differential in the polar coordinates. We shall use

two methods: directly in the polar coordinates and working first in the coordinates x, y and then transforming to r, φ . Directly in polars, we simply have $df = 2rdr$. Now, by the second method we get

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = 2x dx + 2y dy.$$

Plugging $dx = \cos \varphi dr - r \sin \varphi d\varphi$ and $dy = \sin \varphi dr + r \cos \varphi d\varphi$, and multiplying through, after simplification we obtain the same result $df = dr$.

Remark 1.8. The relation between vectors and differentials of functions (considered as linear functions on vectors, at a given point) remain valid in all coordinate systems. The differential df at a point $\mathbf{x}$ is a linear function on vectors attached to $\mathbf{x}$. In particular, for dx^i we have

$$dx^i(\mathbf{e}_j) = dx^i \left(\frac{\partial \mathbf{x}}{\partial x^j} \right) = \frac{\partial x^i}{\partial x^j} = \delta_j^i \quad (52)$$

(the second equality because the value of df on a velocity vector is the derivative of f w.r.t. the parameter on the curve). Hence the differentials of the coordinates dx^i form a basis (in the space of linear functions on vectors) “dual” to the basis $\mathbf{e}_j$.

Problem 1.3. Consider spherical coordinates in $\mathbb{R}^3$. Find the basis $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi$ associated with it (in terms of the standard basis). Find the differentials $dr, d\theta, d\varphi$. Do the same for cylindrical coordinates.

2 Line integrals and 1-forms

We are already acquainted with examples of 1-forms. Let us give a formal definition.

Definition 2.1. A *linear differential form* or, shortly, a *1-form* in an open domain $U \subset \mathbb{R}^n$ is an expression of the form

$$\omega = \omega_1 dx^1 + \dots + \omega_n dx^n,$$

where ω_i are functions. Here $x^1, \dots, x^n$ denote some coordinates in U .

Greek letters like ω, σ , as well as capital Latin letters like A, E , are traditionally used for denoting 1-forms.

Example 2.1. $x dy - y dx, 2dx - (x + z) dy + xy dz$ are examples of 1-forms in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example 2.2. The differential of a function in $\mathbb{R}^n$ is a 1-form:

$$df = \frac{\partial f}{\partial x^1} dx^1 + \dots + \frac{\partial f}{\partial x^n} dx^n.$$

(Notice that not every 1-form is df for some function f . We will see examples later.)

Though Definition 2.1 makes use of some (arbitrary) coordinate system, the notion of a 1-form is independent of coordinates. There are at least two ways to explain this.

Firstly, if we change coordinates, we will obtain again a 1-form (i.e., an expression of the same type).

Example 2.3. Consider a 1-form in $\mathbb{R}^2$ given in the standard coordinates:

$$A = -y \, dx + x \, dy.$$

In the polar coordinates we will have $x = r \cos \varphi$, $y = r \sin \varphi$, hence

$$\begin{aligned} dx &= \cos \varphi \, dr - r \sin \varphi \, d\varphi \\ dy &= \sin \varphi \, dr + r \cos \varphi \, d\varphi. \end{aligned}$$

Substituting into A , we get $A = -r \sin \varphi (\cos \varphi \, dr - r \sin \varphi \, d\varphi) + r \cos \varphi (\sin \varphi \, dr + r \cos \varphi \, d\varphi) = r^2 (\sin^2 \varphi + \cos^2 \varphi) \, d\varphi = r^2 \, d\varphi$. Hence

$$A = r^2 \, d\varphi$$

is the formula for A in the polar coordinates. In particular, we see that this is again a 1-form, a linear combination of the differentials of coordinates with functions as coefficients.

Secondly, in a more conceptual way, we can define a 1-form in a domain U as a linear function on vectors at every point of U :

$$\omega(\mathbf{v}) = \omega_1 v^1 + \dots + \omega_n v^n, \quad (53)$$

if $\mathbf{v} = \sum \mathbf{e}_i v^i$, where $\mathbf{e}_i = \partial \mathbf{x} / \partial x^i$. Recall that the differentials of functions were defined as linear functions on vectors (at every point), and

$$dx^i(\mathbf{e}_j) = dx^i \left(\frac{\partial \mathbf{x}}{\partial x^j} \right) = \delta_j^i \quad (54)$$

at every point $\mathbf{x}$. Remark: if we need to show explicitly the dependence on point, we write $\omega(\mathbf{x})(\mathbf{v})$ (similarly as we did for differentials). There is an alternative notation:

$$\langle \omega, \mathbf{v} \rangle = \omega(\mathbf{v}) = \omega_1 v^1 + \dots + \omega_n v^n, \quad (55)$$

which is sometimes more convenient. (Notice angle brackets; do not confuse it with a scalar product of vectors, which is defined for an Euclidean space.)

Example 2.4. At the point $\mathbf{x} = (1, 2, -1) \in \mathbb{R}^3$ (we are using the standard coordinates) calculate $\langle \omega, \mathbf{v} \rangle$ if $\omega = x dx + y dy + z dz$ and $\mathbf{v} = 3\mathbf{e}_1 - 5\mathbf{e}_3$. We have

$$\langle \omega(\mathbf{x}), \mathbf{v} \rangle = \langle dx + 2dy - dz, 3\mathbf{e}_1 - 5\mathbf{e}_3 \rangle = 3 - (-5) = 8.$$

The main purpose for which we need 1-forms is integration.

Suppose we are given a path, i.e., a parametrized curve in $\mathbb{R}^n$. Denote it $\gamma: x^i = x^i(t), t \in [a, b] \subset \mathbb{R}$. Consider a 1-form ω .

Definition 2.2. The *integral* of ω over γ is

$$\begin{aligned} \int_{\gamma} \omega &= \int_{\gamma} \omega_1 dx^1 + \dots + \omega_n dx^n := \\ \int_a^b \langle \omega(\mathbf{x}(t)), \dot{\mathbf{x}}(t) \rangle dt &= \int_a^b \left(\omega_1(\mathbf{x}(t)) \frac{dx^1}{dt} + \dots + \omega_n(\mathbf{x}(t)) \frac{dx^n}{dt} \right) dt. \end{aligned} \quad (56)$$

Integrals of 1-forms are also called *line integrals*. For a line integral we need two ingredients: a path of integration and a 1-form. The integral depends on both.

Example 2.5. Consider a “constant” 1-form $E = 2dx - 3dy$ in $\mathbb{R}^2$. Let γ be the following path: $x = t, y = 1 - t$, where $t \in [0, 1]$. (It represents a straight line segment $[PQ]$ where $P = (0, 1), Q = (1, 0)$.) To calculate $\int_{\gamma} E$, we first find the velocity vector: $\dot{\mathbf{x}} = (1, -1) = \mathbf{e}_1 - \mathbf{e}_2$ (constant, in this case). Next, we take the value of E on $\dot{\mathbf{x}}$: $\langle E, \dot{\mathbf{x}} \rangle = \langle 2dx - 3dy, \mathbf{e}_1 - \mathbf{e}_2 \rangle = 2 - 3(-1) = 5$. Hence,

$$\int_{\gamma} E = \int_0^1 \langle E, \dot{\mathbf{x}} \rangle dt = \int_0^1 5 dt = 5.$$

Remark 2.1. A practical way of calculating line integrals is based on the following shortcut: the expression $\langle \omega(\mathbf{x}(t)), \dot{\mathbf{x}}(t) \rangle dt$ is simply a 1-form on $[a, b] \subset \mathbb{R}$ obtained from ω by substituting $x^i = x^i(t)$ as functions of t given by the path γ . We have to substitute both in the arguments of $\omega_i(\mathbf{x})$ and in the differentials dx^i expanding them as the differentials of functions of t . The resulting 1-form on $[a, b]$ depends on both ω and γ , and is denoted $\gamma^*\omega$:

$$\gamma^*\omega = \left(\sum \omega_i(\mathbf{x}(t)) \frac{dx^i}{dt}(t) \right) dt, \quad (57)$$

so that

$$\int_{\gamma} \omega = \int_a^b \gamma^*\omega. \quad (58)$$

Example 2.6. Find the integral of the 1-form $A = x dy - y dx$ over the path $\gamma: x = t, y = t^2, t \in [0, 2]$. Considering x, y as functions of t (given by the path γ), we can calculate their differentials: $dx = dt, dy = 2t dt$. Hence

$$\int_{\gamma} A = \int_0^2 \gamma^*A = \int_0^2 (t(2t dt) - t^2 dt) = \int_0^2 t^2 dt = \left. \frac{t^3}{3} \right|_0^2 = \frac{8}{3}$$

Definition 2.3. An *orientation* of a path $\gamma: \mathbf{x} = \mathbf{x}(t)$ is given by the direction of the velocity vector $\dot{\mathbf{x}}$.

Suppose we change parametrization of a path γ . That means we consider a new path $\gamma': [a', b'] \rightarrow \mathbb{R}^n$ obtained from γ by a substitution $t = t(t')$. We use t' to denote the new parameter. Let us assume that $dt/dt' \neq 0$, i.e., the function $t' \mapsto t = t(t')$ is monotonous. If it increases, then the velocity vectors $d\mathbf{x}/dt$ and $d\mathbf{x}/dt'$ have the same direction; if it decreases, they have the opposite directions. Recall that

$$\frac{d\mathbf{x}}{dt'} = \frac{d\mathbf{x}}{dt} \cdot \frac{dt}{dt'}. \quad (59)$$

The most important statement concerning line integrals is the following theorem.

Theorem 2.1. For arbitrary 1-form ω and path γ , the integral $\int_{\gamma} \omega$ does not change if we change parametrization of γ provided the orientation remains the same.

Proof. Consider $\langle \omega(\mathbf{x}(t)), \frac{d\mathbf{x}}{dt} \rangle$ and $\langle \omega(\mathbf{x}(t(t'))), \frac{d\mathbf{x}}{dt'} \rangle$. As

$$\left\langle \omega(\mathbf{x}(t(t'))), \frac{d\mathbf{x}}{dt'} \right\rangle = \left\langle \omega(\mathbf{x}(t(t'))), \frac{d\mathbf{x}}{dt} \right\rangle \cdot \frac{dt}{dt'},$$

we can use the familiar formula

$$\int_a^b f(t) dt = \int_{a'}^{b'} f(t(t')) \cdot \frac{dt}{dt'} dt'$$

valid if $\frac{dt}{dt'} > 0$, and the statement immediately follows. $\square$

If we change orientation to the opposite, then the integral changes sign. This corresponds to the formula

$$\int_a^b f(t) dt = - \int_b^a f(t) dt$$

in the calculus of a single variable.

Independence of parametrization allows us to define line integrals over more general objects. We can consider integrals over any “contours” consisting of pieces which can be represented by parametrized curves; such contours can have “angles” and not necessarily be connected. We simply add integrals over pieces. All that we need to calculate the integral of a 1-form over a contour is an orientation of the contour, i.e., an orientation for every piece that can be represented by a parametrized curve.

Example 2.7. Consider in $\mathbb{R}^2$ a contour $ABCD$ consisting of the segments $[AB]$, $[BC]$, $[CD]$, where $A = (-2, 0)$, $B = (-2, 4)$, $C = (2, 4)$, $D = (2, 0)$ (this is an upper part of the boundary of the square $ABCD$). The orientation is given by the order of vertices $ABCD$. Calculate $\int_{ABCD} \omega$ for $\omega = (x + y) dx + y dy$. The integral is the sum of the three integrals over the segments $[AB]$, $[BC]$ and $[CD]$. As parameters we can take y for $[AB]$ and $[CD]$, and x for $[BC]$. We have

$$\begin{aligned}\int_{[AB]} \omega &= \int_0^4 y dy = \frac{y^2}{2} \Big|_0^4 = 8 \\ \int_{[BC]} \omega &= \int_{-2}^2 (x + 4) dx = \frac{(x + 4)^2}{2} \Big|_{-2}^2 = 18 - 2 = 16 \\ \int_{[CD]} \omega &= \int_4^0 y dy = -8\end{aligned}$$

Hence

$$\int_{ABCD} \omega = \int_{[AB]} \omega + \int_{[BC]} \omega + \int_{[CD]} \omega = 8 + 16 - 8 = 16.$$

Notice that the integrals over vertical segments cancel each other.

Example 2.8. Calculate the integral of the form $\omega = dz$ over the perimeter of the triangle ABC in $\mathbb{R}^3$ (orientation is given by the order of vertices), if $A = (1, 0, 0)$, $B = (0, 2, 0)$, $C = (0, 0, 3)$. We can parameterize the sides of the triangle as follows:

$$\begin{aligned}[AB]: & x = 1 - t, y = 2t, z = 0, \quad t \in [0, 1] \\ [BC]: & x = 0, y = 2 - 2t, z = 3t, \quad t \in [0, 1] \\ [CA]: & x = t, y = 0, z = -3t, \quad t \in [0, 1].\end{aligned}$$

Hence

$$\int_{[AB]} dz = 0, \quad \int_{[BC]} dz = \int_0^1 3dt = 3, \quad \int_{[CA]} dz = \int_0^1 (-3dt) = -3;$$

and the integral in question is $I = 0 + 3 - 3 = 0$.

3 Algebra of forms

3.1 Jacobian

Recall the formula for the transformation of variables in the double integral:

$$\int_D f(x, y) dx dy = \pm \int_D f(x(u, v), y(u, v)) J(u, v) du dv \quad (60)$$

where

$$J = \frac{D(x, y)}{D(u, v)} = \det \frac{\partial(x, y)}{\partial(u, v)} \quad (61)$$

is called the *Jacobian* of the transformation of variables. Here

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} \quad (62)$$

denotes the matrix of partial derivatives. The sign $\pm$ in (60) is the sign of the Jacobian.

Example 3.1. For polar coordinates r, φ where $x = r \cos \varphi, y = r \sin \varphi$, we have $dx = \cos \varphi dr - r \sin \varphi d\varphi$, $dy = \sin \varphi dr + r \cos \varphi d\varphi$, hence

$$\frac{D(x, y)}{D(r, \varphi)} = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r. \quad (63)$$

Thus we can write $dx dy = r dr d\varphi$.

Problem 3.1. Calculate the area of a disk of radius R in two ways: using the standard coordinates and using polar coordinates. By “area” we mean $\int_D dx dy$. (Here $D = \{(x, y) \mid -\sqrt{R^2 - x^2} \leq y \leq \sqrt{R^2 - x^2}, -R \leq x \leq R\} = \{(r, \varphi) \mid 0 \leq r \leq R, 0 \leq \varphi \leq 2\pi\}$.)

We can ask ourselves the following question: is there a way of getting the formula with the Jacobian by multiplying the formulae for the differentials?

Or, to put it differently: is it possible to understand $dx dy$ as an actual product of the differentials dx and dy so that the formula like $dx dy = r dr d\varphi$ in the above example comes naturally?

The answer is “yes”. The rules that we have to adopt for the multiplication of $dr, d\varphi$ are as follows:

$$dr dr = 0, dr d\varphi = -d\varphi dr, d\varphi d\varphi = 0. \quad (64)$$

Indeed, if we calculate according to these rules, we get: $dx dy = (\cos \varphi dr - r \sin \varphi d\varphi)(\sin \varphi dr + r \cos \varphi d\varphi) = \cos \varphi \sin \varphi dr dr + r \cos^2 \varphi dr d\varphi - r \sin^2 \varphi d\varphi dr - r^2 \sin \varphi \cos \varphi d\varphi d\varphi = r \cos^2 \varphi dr d\varphi - r \sin^2 \varphi (-dr d\varphi) = r dr d\varphi$, as required. (We also assumed the usual distributivity w.r.t. addition.) These are the only rules that lead to the correct answer.

Notice that as a corollary we get similar rules for the products of dx and dy :

$$dx dx = 0, dx dy = -dy dx, dy dy = 0 \quad (65)$$

(check!).

More generally, we can check that in this way we will get the general formula for arbitrary changes of variables in $\mathbb{R}^2$: if $x = x(u, v)$, $y = y(u, v)$, then

$$dx = \frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv \quad (66)$$

$$dy = \frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv, \quad (67)$$

and we get

$$\begin{aligned} dx dy &= \left(\frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv \right) \left(\frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv \right) = \\ &= \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} du dv + \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} dv du = \left(\frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial x}{\partial v} \frac{\partial y}{\partial u} \right) du dv = J du dv \end{aligned}$$

(we used $du dv = -dv du$; there are no terms with $du du = 0$ and $dv dv = 0$).

The multiplication that we have just introduced is called the “exterior multiplication”. The characteristic feature of the “exterior” rules is that the product of two differentials changes sign if we change the order. In particular, the exterior product of any differential with itself must vanish. Notice that a product like $dx dy$ is a new object compared to dx and dy : it does not equal any of 1-forms, and it is the kind of expression that stands under the sign of integral when we integrate over a two-dimensional region.

Now, if we consider $\mathbb{R}^n$, a similar formula with the Jacobian is valid

$$\int_D f dx^1 \dots dx^n = \pm \int_D f J du^1 \dots du^n \quad (68)$$

where

$$J = \frac{D(x^1, \dots, x^n)}{D(u^1, \dots, u^n)} = \det \frac{\partial(x^1, \dots, x^n)}{\partial(u^1, \dots, u^n)}. \quad (69)$$

How to handle it? Can we understand it in the same way as the formula in two dimensions? It turns out that all we have to do is to extend the exterior multiplication from just two to an arbitrary number of factors. We discuss it in the next section.

3.2 Rules of exterior multiplication

Definition of forms in $\mathbb{R}^n$. Definition of the exterior product. Examples.

Example: translation; linear change; $x + xy$.

Effect of maps.

Jacobian obtained from n -forms.

Remark 3.1. As well as dx^i as a linear function on vectors gives the i -th coordinate: $dx^i(\mathbf{h}) = h^i$, the exterior product $dx^i dx^j$ can be understood as a function on pairs of vectors giving the determinant

$$dx^i dx^j(\mathbf{h}_1, \mathbf{h}_2) = \begin{vmatrix} h_1^i & h_2^i \\ h_1^j & h_2^j \end{vmatrix} = h_1^i h_2^j - h_1^j h_2^i, \quad (70)$$

and similarly for the triple exterior products like $dx^i dx^j dx^k$, etc. We will not pursue it further.

4 Exterior derivative

Want to get $d: \Omega^k \rightarrow \Omega^{k+1}$

4.1 Dependence of line integrals on paths

Consider paths in $\mathbb{R}^n$

Theorem 4.1. *The integral of df does not depend on path (provided the endpoints are fixed).*

Closed paths that are boundaries: additivity of integral. Small path.

Let us explore how $\int_\gamma \omega$ depend on path in general. For this we shall consider a “small” closed path and calculate the integral over it....

d for 1-forms

4.2 Exterior derivative: construction

Consider differential forms defined in some open domain $U \subset \mathbb{R}^n$. In the sequel we omit further references to U and simply write Ω^k for $\Omega^k(U)$. An axiomatic definition of d is given by the following theorem:

Theorem 4.2. *There is a unique operation $d: \Omega^k \rightarrow \Omega^{k+1}$ for all $k = 0, 1, \dots, n$ possessing the following properties:*

$$d(a\omega + b\sigma) = a d\omega + b d\sigma \quad (71)$$

for any k -forms ω and σ (where a, b are constants);

$$d(\omega\sigma) = (d\omega)\sigma + (-1)^k \omega(d\sigma) \quad (72)$$

for any k -form ω and l -form σ ; on functions

$$df = \sum \frac{\partial f}{\partial x^i} dx^i \quad (73)$$

is the usual differential, and

$$d(df) = 0 \quad (74)$$

for any function f .

Proof. Let us assume that an operator d satisfying these properties exists. By induction we deduce that $d(dx^{i_1} \dots dx^{i_k}) = 0$. Hence for an arbitrary k -form $\omega = \sum_{i_1 < \dots < i_k} \omega_{i_1 \dots i_k} dx^{i_1} \dots dx^{i_k}$ we arrive at the formula

$$d\omega = \sum_{i_1 < \dots < i_k} d\omega_{i_1 \dots i_k} dx^{i_1} \dots dx^{i_k}.$$

(This establishes the uniqueness part of the theorem.) Now we take this formula as a definition. By a direct check we can show that it will satisfy all required properties. For example, to show that $d(df) = 0$ for an arbitrary function f , we apply the above formula to df and obtain

$$\begin{aligned} d(df) &= d\left(\sum \frac{\partial f}{\partial x^i} dx^i\right) = \sum \frac{\partial^2 f}{\partial x^j \partial x^i} dx^j dx^i = \\ &= -\sum \frac{\partial^2 f}{\partial x^i \partial x^j} dx^i dx^j = -\sum \frac{\partial^2 f}{\partial x^j \partial x^i} dx^j dx^i = -d(df) \end{aligned}$$

(we swapped the partial derivatives using their commutativity and the differentials in the exterior product $dx^j dx^i$ getting the minus sign, and then renamed the summation indices). Thus $d(df) = 0$, as required. (This establishes the existence part of the theorem.) $\square$

The formula

$$d\omega = \sum_{i_1 < \dots < i_k} d\omega_{i_1 \dots i_k} dx^{i_1} \dots dx^{i_k}$$

is the working formula for the calculation of d . In many cases, though, it is easier to calculate directly using the basic properties of d . On the other hand, one can deduce “general” explicit expressions in particular cases (like 1-forms).

4.3 Main properties and examples of calculation

Example 4.1. Suppose ω is a $(n-1)$ -form in $\mathbb{R}^n$. We can write it as

$$\omega = \sum \omega^i dx^1 \dots dx^{i-1} dx^{i+1} \dots dx^n. \quad (75)$$

Then we easily obtain

$$d\omega = \sum \frac{\partial \omega^i}{\partial x^i} dx^1 \dots dx^n. \quad (76)$$

Exact and closed forms

Theorem 4.3. $d \circ F^* = F^* \circ d$

5 Stokes's theorem

The Stokes theorem is one of the most fundamental results that you will learn from the university mathematics course. It has the form of the equality

$$\int_C d\omega = \int_{\partial C} \omega,$$

where $\omega \in \Omega^{k-1}(U)$ is a $(k-1)$ -form, C a “ k -dimensional contour” (to be defined) and ∂C the boundary of this contour (to be defined, as well).

There are two cases already familiar.

The Newton–Leibniz formula (or the “main theorem of integral calculus”) is for $k = 1$. Then $k - 1 = 0$, and $\omega = f$ is just a function. If we take a path $\gamma: [0, 1] \rightarrow U$ as a contour, then

$$\int_{\gamma} df = f(Q) - f(P) \quad (77)$$

where $P = \gamma(0)$, $Q = \gamma(1)$. The r.h.s. should be considered as an “integral of a 0-form” over the boundary of the path γ , which is a formal linear combination of its endpoints: $Q - P$; i.e., we treat $f(Q) - f(P)$ as the “integral” of f over $Q - P$. The integral of a function (as a 0-form) over a point is defined as the value at this point. It is essential that the path has the orientation (direction) “from P to Q ”.

The Green formula in the plane is the case $k = 2$ (hence $k - 1 = 1$) and $n = 2$. Then $\omega = \omega_1 dx^1 + \omega_2 dx^2$. For a bounded domain $D \subset \mathbb{R}^2$ with the boundary being a closed curve C , we have the formula

$$\int_D \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2} \right) dx^1 dx^2 = \int_C (\omega_1 dx^1 + \omega_2 dx^2). \quad (78)$$

Under the integral at the l.h.s. stands precisely the differential $d\omega$. Again, it is important that the domain D and its boundary C are given coherent orientations.

In this section we shall define all notions involved in Stokes's theorem, discuss its applications and give an idea of a proof.

5.1 Integration of k -forms

We are going to define integrals like $\int_C \omega$ where $\omega \in \Omega(U)$, $U \subset \mathbb{R}^n$ is an open domain, and C is a “ k -dimensional contour”. (Such “contours” will be called “chains”.) This will be achieved in several steps.

Step 1. The “contours” are supposed to be oriented. So we have to introduce the notion of an orientation. Let $D \subset \mathbb{R}^k$ be some bounded domain. What is an “orientation” of D ? Consider examples.

$k = 0$. D is just a point. An orientation of a point is simply a sign (plus or minus) assigned to this point. For example, in the Newton–Leibniz formula above we meet the expression $f(Q) - f(P)$; it should be treated as an integral of a 0-form f over the formal difference $Q - P$, i.e., over the points P and Q where to P we assign $+1$ and to Q we assign -1 .

$k = 1$. $D = [a, b]$ is a segment. An orientation is a (choice of) direction: from a to b or from b to a .

$k = 2$. We have D a domain in $\mathbb{R}^2$. An orientation here is a sense of rotation, which we may describe as “counterclockwise” or “clockwise”. It is important to realize that these terms have no absolute meaning: if we look at a sketch of our domain from the other side of the paper, what was “counterclockwise” will be “clockwise”, and vice versa. The labels by which we distinguish the two possible orientations of D are relative; important is that there are exactly two of them. (We assume that D is connected; otherwise each connected component can be assigned one of the two orientations independently.) How one can specify an orientation? A general method is to choose a basis of vectors $\mathbf{e}_1, \mathbf{e}_2$ (at a point of D) and by “moving” it to other points of D get a basis at all points continuously depending on a point. Then the sense of rotation at every point will be given as from $\mathbf{e}_1$ to $\mathbf{e}_2$. It is clear that if we slightly deform the basis (move the vectors $\mathbf{e}_1, \mathbf{e}_2$ at each point slightly) or multiply it by a matrix with a positive determinant, the sense of rotation defined by $\mathbf{e}_1, \mathbf{e}_2$ will not change. In particular, if we have a coordinate system x^1, x^2 in D and take the basis $\mathbf{e}_1 = \partial \mathbf{x} / \partial x^1, \mathbf{e}_2 = \partial \mathbf{x} / \partial x^2$, then any two coordinate systems define the same orientation if and only if the Jacobian of the transformation of coordinates is positive, and they define the opposite orientations if the Jacobian is negative.

Example 5.1. Consider in $\mathbb{R}^2$ the standard coordinates x, y and the polar coordinates r, φ so that $x = r \cos \varphi, y = r \sin \varphi$. Then

$$J = \frac{D(x, y)}{D(r, \varphi)} = \begin{vmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & r \cos \varphi \end{vmatrix} = r > 0, \quad (79)$$

hence the coordinate systems x, y and r, φ define the same orientation in the plane. The order of coordinates is very important: if we consider φ, r instead (the φ coordinate considered as the first and the r coordinate considered as the second), then the columns of the determinant in (79) will be swapped and the Jacobian will change sign. In general, if we swap two coordinates, then the orientation changes to the opposite. Also, if we change φ to $\varphi' = -\varphi$ (i.e., begin to count the polar angles “clockwise” instead of “counterclockwise”), then the second column will change sign, again giving a negative Jacobian.

$k > 2$. In this general case we define an orientation in $D \subset \mathbb{R}^k$ using a basis $\mathbf{e}_1, \dots, \mathbf{e}_k$ at every point of D (we assume that the vectors of the basis

smoothly depend on point). Two bases $\mathbf{e}_1, \dots, \mathbf{e}_k$ and $\mathbf{g}_1, \dots, \mathbf{g}_k$ are said to be *equivalent* or *defining the same orientation* if and only if $\mathbf{g}_i = \sum_{j=1}^k \mathbf{e}_j A_{ji}$ with $\det A > 0$. Notice that every two bases in a vector space differ by a linear transformation with an invertible matrix. The point is that the matrix should have a positive determinant for the bases in question to define the same orientation. Notice that the matrix A is invertible, hence $\det A$ is never zero. It follows that if we slightly perturb the basis, i.e., take A close to the identity matrix E , or consider a continuous family of transformations starting at the identity, then the determinant cannot change sign (as $\det E = 1$ and passing through 0 is impossible), so the orientation will not change.

In particular, it is possible to take the bases corresponding to coordinate systems in D . Then any two coordinate systems, say, $x^1, \dots, x^k$ and $y^1, \dots, y^k$ define the same orientation in D if

$$J = \frac{D(x^1, \dots, x^k)}{D(y^1, \dots, y^k)} = \begin{vmatrix} \frac{\partial x^1}{\partial y^1} & \cdots & \frac{\partial x^1}{\partial y^k} \\ \vdots & \ddots & \vdots \\ \frac{\partial x^k}{\partial y^1} & \cdots & \frac{\partial x^k}{\partial y^k} \end{vmatrix} > 0, \quad (80)$$

and they define the opposite orientations if the Jacobian is negative. (This definition applied to $k = 1$ and $k = 2$ embraces the cases considered separately above. It turns out that only the case $k = 0$ requires an *ad hoc* definition.)

Problem 5.1. For spherical coordinates in $\mathbb{R}^3$ find out which order of coordinates: r, φ, θ or r, θ, φ gives the same orientation as x, y, z .

Step 2. Now we shall define the integral of an n -form over a bounded domain $D \subset \mathbb{R}^n$. Consider $\omega \in \Omega^n(\mathbb{R}^n)$. Let $x^1, \dots, x^n$ be some coordinates in D . Since there is just one independent n -fold product of the differentials dx^i here, the form ω has a very simple appearance:

$$\omega = f(x^1, \dots, x^n) dx^1 \dots dx^n, \quad (81)$$

where f is a function of the coordinates. We define the *integral of an n -form ω over an n -dimensional domain D* as

$$\int_D \omega = \int_D f(x^1, \dots, x^n) dx^1 \dots dx^n, \quad (82)$$

where at the r.h.s. stands the usual multiple integral as defined in calculus courses (via partitions and limits of integral sums). Let us recall the theorem about the change of variables in multiple integrals. If we introduce “new” coordinates $y^1, \dots, y^n$ in the domain D , then

$$\begin{aligned} \int_D f(x^1, \dots, x^n) dx^1 \dots dx^n = \\ \pm \int_D f(x^1(y^1, \dots, y^n), \dots, x^n(y^1, \dots, y^n)) J dy^1 \dots dy^n. \end{aligned} \quad (83)$$

Here $J = J(y^1, \dots, y^n)$ is the Jacobian $D(x^1, \dots, x^n)/D(y^1, \dots, y^n)$. Forgetting for a moment about the sign $\pm$ in (83), we recognize the l.h.s. as $\int_D \omega$ (defined using the coordinates $x^1, \dots, x^n$) and the expression under the integral in the r.h.s. as exactly the form ω written in the coordinates $y^1, \dots, y^n$. Indeed, the rules of the exterior multiplication are devised precisely to give $dx^1 \dots dx^n = J dy^1 \dots dy^n$ when x^i are expressed as functions of y^j (see Section 3.1). It follows that the integral in the r.h.s. is the integral of ω calculated in the coordinates y^j . Hence, up to a sign, we see that the integral $\int_D \omega$ does not depend on a choice of coordinates in D . Now, what is the meaning of the sign? Notice that $\pm = \text{sign } J$. We obtain the following theorem.

Theorem 5.1. *The integral of an n -form ω over $D \subset \mathbb{R}^n$ does not depend on a choice of coordinates in D provided the orientation is fixed. If we change orientation, the integral changes sign.* $\square$

Example 5.2. Let $I^2 = \{(x, y) \mid 0 \leq x, y \leq 1\}$ be the unit square in $\mathbb{R}^2$. Let an orientation be given by the coordinates x, y . Then

$$\int_{I^2} dx dy = - \int_{I^2} dy dx = 1. \quad (84)$$

Define an *oriented domain* as a domain D with a chosen orientation. Let $-D$ denote the same domain with the opposite orientation. Hence, we can integrate n -forms over oriented domains in $\mathbb{R}^n$, and

$$\int_{-D} \omega = - \int_D \omega. \quad (85)$$

In the above we assumed that $n > 0$. For $n = 0$, we define the integral of a 0-form in $\mathbb{R}^0$, which just a number ($\mathbb{R}^0$ is a single point!), to be this number if the point is taken with $+$ or minus this number if the point is taken with $-$ (see above on an orientation for $k = 0$).

Step 3. Now we want to learn how to integrate k -forms in $\mathbb{R}^n$. We have to explain first over which objects forms will be integrated. Bear in mind the analogy with $k = 1$. We define a k -path or a k -dimensional path in $\mathbb{R}^n$ or in $U \subset \mathbb{R}^n$ (an open domain) as a smooth map $\Gamma: D \rightarrow U$, where $D \subset \mathbb{R}^k$ is a bounded domain. (Recall that for $k = 1$, a path or a “1-path” is a map $[a, b] \rightarrow U$.) A “parametrization” of a k -path Γ is a choice of coordinates in D . For any $\omega \in \Omega^k(U)$ define the *integral of a k -form over a k -path*

$$\int_{\Gamma} \omega := \int_D \Gamma^* \omega. \quad (86)$$

Here $\Gamma^*: \Omega^k(U) \rightarrow \Omega^k(D)$ is the pull-back map. The integral at the r.h.s. is the integral of a k -form over a bounded domain in $\mathbb{R}^k$. It does not depend

on a choice of coordinates as long as we do not change the orientation. An orientation of D will be called an *orientation* of the k -path Γ . It follows that the integral of k -forms is well-defined on oriented k -paths. If we agree to denote by Γ an oriented k -path and by $-\Gamma$ the same k -path with the opposite orientation, then we have

$$\int_{-\Gamma} \omega = - \int_{\Gamma} \omega. \quad (87)$$

Since we can integrate over k -paths so that the integral depends only on orientation and not a parametrization, we can extend the integral to any objects that can be “cut” into pieces representable by k -paths, — provided the orientations on the pieces are fixed. Following the 1-dimensional example, we define a *k-chain* (or a *k-dimensional chain*) in U as a formal linear combination of oriented k -paths:

$$C = a_1\Gamma_1 + a_2\Gamma_2 + \dots + a_r\Gamma_r, \quad (88)$$

where $a_i \in \mathbb{R}$ and Γ_i are k -dimensional paths with chosen orientations. When we take linear combinations we agree that $-(-\Gamma) = \Gamma$, where the minus sign in the brackets means taking the opposite orientation (so that the “formal” multiplication by -1 coincides with changing the orientation). We define the *integral of a k -form in $U \subset \mathbb{R}^n$ over a k -chain* as the sum of integrals over k -paths:

$$\int_C \omega = \int_{a_1\Gamma_1 + \dots + a_r\Gamma_r} \omega := a_1 \int_{\Gamma_1} \omega + \dots + a_r \int_{\Gamma_r} \omega. \quad (89)$$

Remark 5.1. The standard definitions of chains are more restrictive than the one given above. We have not specified bounded domains $D \subset \mathbb{R}^k$ that are allowed in k -paths. Such flexibility is convenient for practical calculations, but makes it difficult to advance with general theorems. For theoretical purposes, as the experience shows, it is better to consider chains made of k -paths defined as maps $\Gamma: D \rightarrow U$ where D is a standard domain like the unit k -cube (cubical chains) or the standard k -simplex (simplicial chains).

Remark 5.2. To integrate over a k -dimensional surface, like a sphere, we have to represent it by a chain. First of all, a surface should be oriented, i.e., an orientation should be chosen at every tangent plane by a basis continuously depending on a point. If it is not possible, then the integral cannot be defined. An oriented surface can be cut into pieces representable by oriented k -paths, so that the whole surface can be replaced by their sum (a chain). The integral over an oriented surface is defined as the integral over this chain. Clearly, the integral will not depend on a particular cutting into pieces, i.e., a representation of the surface by a chain.

Example 5.3. Calculate the integral of the 2-form $x dx dy + (2 + 3y) dx dz$ over the 2-path $\Gamma: D \rightarrow \mathbb{R}^3$, $\Gamma: (u, v) \mapsto (u, v, u^2 + v^2)$, and $D: 0 \leq u, v \leq 1$. We find $\Gamma^*\omega = u du dv + (2 + 3v) du d(u^2 + v^2) = (u + 2 + 6v^2) du dv$. Hence

$$\int_{\Gamma} \omega = \int_D (u + 2 + 6v^2) du dv = \int_0^1 \int_0^1 (u + 2 + 6v^2) du dv = \int_0^1 (u + 2 + 2) du = 4 \frac{1}{2}.$$

Example 5.4. Calculate the integral of the 2-form $\omega = z dx dy$ over the the unit sphere with center $O = (0, 0, 0)$ in $\mathbb{R}^3$. Denote this surface S^2 . An orientation is fixed by the basis $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$ at the point $(0, 0, 1)$ (the “north pole”). At all other points of the sphere we can get a basis defining the orientation by dragging to these points the given basis at $(0, 0, 1)$ (so that it remains tangent to the sphere). We have to represent the sphere by a chain. One way of doing this is to cut it into the upper hemisphere and the lower hemisphere. Consider the 2-paths Γ_+ and Γ_- , where $\Gamma_+: D \rightarrow \mathbb{R}^3$, $(u, v) \mapsto (u, v, \sqrt{1 - u^2 - v^2})$, and $\Gamma_-: D \rightarrow \mathbb{R}^3$, $(u, v) \mapsto (u, v, -\sqrt{1 - u^2 - v^2})$. Here $D = \{u^2 + v^2 \leq 1\}$ (for both 2-paths). Now, what about the orientation? One can see that at the south pole $(0, 0, -1)$ the given orientation of the sphere S^2 is opposite to that specified by $\mathbf{e}_1, \mathbf{e}_2$ at this point (check!). Hence for Γ_+ we should take the orientation by the coordinates u, v , and for Γ_- the opposite orientation, so the sphere is represented by the 2-chain $C = \Gamma_+ - \Gamma_-$. We have

$$\begin{aligned} \int_{S^2} \omega &= \int_C \omega = \int_{\Gamma_+} \omega - \int_{\Gamma_-} \omega = \int_D \Gamma_+^* \omega - \int_D \Gamma_-^* \omega = \\ &= \int_D \sqrt{1 - u^2 - v^2} du dv - \int_D (-\sqrt{1 - u^2 - v^2} du dv) = \\ &= 2 \int_D \sqrt{1 - u^2 - v^2} du dv \end{aligned}$$

(we skip the calculation of the remaining double integral).

Problem 1. Calculate the integral $\int_{S^2} \omega$ using spherical coordinates (i.e., representing the sphere by the chain consisting of a single 2-path: $x = \cos \varphi \sin \theta$, $y = \sin \varphi \sin \theta$, $z = \cos \theta$, $0 \leq \varphi \leq 2\pi$, $0 \leq \theta \leq \pi$) and check that you will get the same answer.

5.2 Stokes's theorem: statement and examples

Theorem 5.2 (Stokes's theorem for chains). *Let $U \subset \mathbb{R}^n$ be an open domain. For every $\omega \in \Omega^{k-1}(U)$ and every k -chain C in U ,*

$$\int_{\partial C} \omega = \int_C d\omega, \quad (90)$$

where ∂C is the boundary of C .

The only notion that we have not discussed yet is the “boundary of a chain”.

Remark 5.3. The theorem remains true if we replace $U \subset \mathbb{R}^n$ by any “smooth n -dimensional manifold”. Smooth manifolds are sets that locally look like open domains of $\mathbb{R}^n$. All calculus, including the theory of differential forms, can be extended to smooth manifolds.

Remark 5.4. In the statement of the theorem, the chain C can be replaced by an “oriented submanifold with boundary”, e.g., a closed disk.

The *boundary* of a k -chain C is a $(k - 1)$ -chain defined as follows. If $C = a_1\Gamma_1 + \dots + a_r\Gamma_r$, we set $\partial C = a_1\partial\Gamma_1 + \dots + a_r\partial\Gamma_r$. Hence we have to define the chain $\partial\Gamma$ for an oriented k -path Γ . Loosely speaking, if $\Gamma: D \rightarrow U$ and an orientation of D is fixed, then $\partial\Gamma$ is given by the map Γ restricted to ∂D , where ∂D is the boundary of the domain $D \subset \mathbb{R}^k$ in the natural sense, endowed with the orientation compatible with that of D .

Example 5.5. Suppose $D = [a, b] \subset \mathbb{R}$ and the orientation is from a to b . Then the boundary of D consists of a and b , and in view of the Newton–Leibniz formula it is natural to take a with “minus” and b with “plus”. We can consider $\partial[a, b]$ as the chain $b - a$ (formal difference! not the subtraction of numbers).

Example 5.6. Suppose $D \subset \mathbb{R}^2$ is the unit disk with center at the origin oriented by the coordinates x, y . Then ∂D is the unit circle with the orientation counterclockwise.

Example 5.7. Suppose $D \subset \mathbb{R}^2$ is the unit square $\{(x, y) \mid 0 \leq x, y \leq 1\}$ oriented by the coordinates x, y . Then ∂D consists of four segments: $[OA]$, $[AB]$, $[BC]$, $[CO]$ with the orientations from O to A , from A to B , from B to C , from C to O . Here $O = (0, 0)$, $A = (1, 0)$, $B = (1, 1)$, $C = (0, 1)$. We can write $\partial D = [OA] + [AB] + [BC] + [CO]$.

Example 5.8. Suppose $D \subset \mathbb{R}^3$ is the unit ball with center at the origin oriented by the coordinates x, y, z . Then ∂D is the unit sphere with the orientation given by the basis $\mathbf{e}_2, \mathbf{e}_3$ at the point $(1, 0, 0)$. Here $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is the standard basis (corresponding to x, y, z).

Problem 5.2. Show that the same orientation on the sphere is given at the point $(0, 0, 1)$ by the basis $\mathbf{e}_1, \mathbf{e}_2$. (Drag the vectors $\mathbf{e}_2, \mathbf{e}_3$ from $(1, 0, 0)$ to $(0, 0, 1)$ along the meridian and see what you will get there.)

We see that the problem is that in some cases (like for the disk or the ball above), the boundary of D is not naturally defined as a chain, though it may be cut into pieces making up a chain. Here is exactly the point where

restricting by some standard domains like a unit cube has a great advantage (see Remark 5.1).

How to induce an orientation on ∂D by a given orientation for D ? Notice that at the points of the boundary it is always defined the direction “inwards” the domain as well as the opposite direction “outwards” the domain. More precisely, for a vector not tangent to the boundary we can always tell if it points “inwards” or “outwards”.

Example 5.9. For the ball $x^2 + y^2 + z^2 \leq 1$, the basis vector $\mathbf{e}_3$ at the point $(0, 0, 1)$ (the north pole) points outwards and at the point $(0, 0, -1)$ (the south pole) it points inwards.

Definition 5.1. Suppose an orientation of a domain $D \subset \mathbb{R}^k$ is given by some basis $\mathbf{e}_1, \dots, \mathbf{e}_k$. Then the *induced orientation* for ∂D is given by a basis $\mathbf{g}_1, \dots, \mathbf{g}_{k-1}$ (consisting of vectors tangent to ∂D) such that

$$(\mathbf{e}_1, \dots, \mathbf{e}_k) \sim (\mathbf{n}, \mathbf{g}_1, \dots, \mathbf{g}_{k-1}),$$

where $\mathbf{n}$ is any vector pointing outwards. Here $\sim$ means the equivalence of bases, i.e., that they define the same orientation.

Problem 5.3. Check that in the examples above the orientation of the boundary is given by this rule.

Example 5.10. Consider in $\mathbb{R}^2$ a domain between two circles (not necessarily concentric). Check that if the domain is oriented in such a way that at the outer circle the induced orientation is counterclockwise, then at the inner circle it is clockwise, and vice versa.

Theorem 5.2 is valid for any chains where the boundary (with the induced orientation) is understood as explained above. It is also valid for “submanifolds” and their boundaries (as we have already mentioned), and again the orientation on the boundary is given as in Definition 5.1.

Proposition 5.1. *For any chain, $\partial(\partial C) = 0$.*

This is easy to prove for chains based on some particular standard domains, like cubical chains or simplicial chains. A chain C such that $\partial C = 0$ is called a *closed chain* or a *cycle*. An example of a 1-cycle is a closed path. Any oriented surface without boundary, like a sphere, can be cut into pieces making up a cycle. In particular, any *boundary*, i.e., a chain which is the boundary of some other chain is a cycle. It is customary to denote integrals over cycles by $\oint$. In this notation the Stokes theorem reads

$$\oint_{\partial C} \omega = \int_C d\omega. \quad (91)$$

The Stokes theorem has the following immediate corollaries:

Corollary 5.1. *The integral of a closed form over a boundary is zero: if $d\omega = 0$, then*

$$\oint_{\partial C} \omega = 0.$$

The integral of an exact form over any cycle is zero: if $\partial C = 0$, then

$$\int_C d\omega = 0.$$

Corollary 5.2. *If two k -chains, C and C' , bound a $(k+1)$ -chain, i.e., $C - C' = \partial W$ for some $(k+1)$ -chain W , then*

$$\int_C \omega = \int_{C'} \omega$$

for any k -form ω .

Consider some elementary examples.

Example 5.11. Let $\omega = a dx + b dy$ in $\mathbb{R}^2$, where a, b are constants. Then $d\omega = 0$, hence

$$\oint_{\partial D} \omega = 0$$

for every bounded domain $D \subset \mathbb{R}^2$.

Example 5.12. Consider the 1-form

$$\omega = \frac{1}{2}(x dy - y dx) \tag{92}$$

in $\mathbb{R}^2$. Then $d\omega = dx dy$. Hence for any bounded domain D ,

$$\oint_{\partial D} \omega = \int_D dx dy = \text{“area of } D\text{”}.$$

(We shall discuss areas at greater detail in the next section.) The contour in the l.h.s. should be oriented counterclockwise.

The form from the previous example can be re-written in polar coordinates as

$$\omega = \frac{1}{2}(x dy - y dx) = \frac{1}{2}r^2 d\varphi \tag{93}$$

(check!). Hence, integration of $\frac{1}{2}r^2 d\varphi$ over the boundary of a domain gives its area.

Example 5.13. The area of a sector of angle $\Delta\varphi$ of a disk of radius R equals

$$\frac{1}{2} \oint_{\partial D} r^2 d\varphi = \frac{1}{2} R^2 \Delta\varphi \tag{94}$$

(the integrals over the two radii are zero, only the integral over the circular arc gives an input). In particular, the area of the whole disk ($\Delta\varphi = 2\pi$) is πR^2 .

Example 5.14. Consider

$$d \ln r = \frac{dr}{r} = \frac{x dx + y dy}{x^2 + y^2}. \quad (95)$$

Since this form is exact in $\mathbb{R}^2 \setminus \{(0)\}$, the integral of it over any cycle in $\mathbb{R}^2 \setminus \{(0)\}$ vanishes.

Example 5.15. Consider the 1-form

$$d\varphi = d \arctan \frac{y}{x} = \frac{x dy - y dx}{x^2 + y^2}. \quad (96)$$

By Corollary 5.2, the integral of this form is the same for all closed paths that go around the origin once, and equals 2π . In general,

$$\oint_C d\varphi = 2\pi n, \quad n \in \mathbb{Z}, \quad (97)$$

if a cycle C “goes around the origin n times” (example: $x = \cos nt, y = \sin nt, t \in [0, 2\pi]$). If a cycle does not go around the origin, the integral is zero. Notice that n can be negative, and $n = 0$ for a cycle that goes around the origin once and then goes back once (one can show that every such cycle is a boundary).

Example 5.16. Consider in $\mathbb{R}^n$ the following $(n-1)$ -form:

$$\omega = \frac{x^1 dx^2 dx^3 \dots dx^n - x^2 dx^1 dx^3 \dots dx^n + \dots + (-1)^{n-1} x^n dx^1 dx^2 \dots dx^{n-1}}{((x^1)^2 + \dots + (x^n)^2)^{n/2}}. \quad (98)$$

For $n = 3$ it is the form

$$\omega = \frac{x dy dz - y dx dz + z dx dy}{(x^2 + y^2 + z^2)^{3/2}} \quad (99)$$

and for $n = 2$ it is the form $d\varphi$ considered in Example 5.15. One can show that $d\omega = 0$ for all n . On the other hand, if S_R stands for the sphere of radius R with center at the origin $O = (0, \dots, 0)$, then

$$\oint_{S_R} \omega = A_n. \quad (100)$$

where $A_n \neq 0$ is a constant independent of R . Indeed, the restriction of ω to S_R coincides with the restriction of the form $\omega' = R^{-n}(x^1 dx^2 dx^3 \dots dx^n - x^2 dx^1 dx^3 \dots dx^n + \dots + (-1)^{n-1} x^n dx^1 dx^2 \dots dx^{n-1})$ defined at all points of $\mathbb{R}^n$. Since $d\omega' = R^{-n}n dx^1 \dots dx^n$, by the Stokes theorem we get

$$\oint_{S_R} \omega = \oint_{S_R} \omega' = \int_{B_R} R^{-n}n dx^1 \dots dx^n = n R^{-n} \text{vol } B_R = n \text{vol } B_1, \quad (101)$$

where B_R denotes the ball of radius R in $\mathbb{R}^n$.

In particular, the form (98) is a closed but not exact form in $\mathbb{R}^n \setminus O$ (otherwise the integral over any cycle would be zero).

5.3 A proof for a simple case

We shall give a proof of the Stokes theorem for cubical chains. In fact, the statement for this case implies the statement in general, but we will not prove this.

Notice first that by linearity the general Stokes formula for chains can be reduced to the formula for a single k -path:

$$\oint_{\partial\Gamma} \omega = \int_{\Gamma} \omega. \quad (102)$$

Since $\partial\Gamma$ is nothing but the map $\Gamma: D \rightarrow U$ restricted to the boundary of the domain D , we have the l.h.s. in (102) equal to

$$\oint_{\partial D} \Gamma^* \omega$$

and the r.h.s. equal to

$$\int_D \Gamma^* d\omega = \int_D d(\Gamma^* \omega),$$

since $d \circ \mathbf{F}^* = \mathbf{F}^* \circ d$ for any map $\mathbf{F}$ (Theorem 4.3). Hence we have to prove that

$$\oint_{\partial D} \Gamma^* \omega = \int_D d(\Gamma^* \omega) \quad (103)$$

for a $(k-1)$ -form $\Gamma^* \omega$ in $D \subset \mathbb{R}^k$.

We can conclude that to prove Stokes's theorem it would suffice to prove it just for a bounded domain in $\mathbb{R}^n$ and $(n-1)$ -forms in it:

$$\oint_{\partial D} \omega = \int_D d\omega. \quad (104)$$

This is one of the strengths of the theory of forms — due to pullbacks and the fact that d commutes with pullbacks. (In obsolete expositions not making use of forms, proofs of statements like Stokes's theorem are much harder exactly because of the lack of pullbacks.) To make (104) more explicit, we can write ω as $\sum \omega^i dx^1 \dots dx^{i-1} dx^{i+1} \dots dx^n$. Then we can use

Remark 5.5. There is a problem of singling out a class of domains D in $\mathbb{R}^n$ for which the Stokes theorem is valid. We do not want to do it precisely. Instead let us just mention that this class should contain all natural examples, like cubes, balls, and any bounded domains specified by a finite set of inequalities $f_i(\mathbf{x}) \geq 0$ in $\mathbb{R}^n$. A good example is the domain between two graphs of functions $x^n = f(x^1, \dots, x^{n-1})$ in $\mathbb{R}^n$ where the argument $(x^1, \dots, x^{n-1})$ runs over a similar type region in $\mathbb{R}^{n-1}$. In all such cases it is possible to define clearly ∂D , the induced orientation for it and the integral

of any $(n-1)$ -form. It seems intuitively clear, — in fact, this is a statement of a topological nature proved in the so-called homology theory, — that every D as above can be cut into “cubical” pieces, i.e., that can be made cubes by changes of variables, so that the sum of the boundaries of all cubes gives a chain representing the boundary of D , with the induced orientation, and all faces that are not on ∂D enter the formula with opposite orientations, hence are cancelled.

Example 5.17. Consider a square in $\mathbb{R}^2$ oriented by coordinates x, y . Cut it into four smaller squares and consider them with the same orientation. Take the boundary of each of the smaller squares (with the induced orientation), and check that all “internal” sides appear twice with the opposite orientations, hence in the sum they cancel, leaving only the sides subdividing the sides of the original square, with the correct orientations.

Now we shall prove the Stokes formula for a cube. It directly implies the theorem for all cubical chains. From the above remark it also follows that it implies the theorem for all other chains (and actually for manifolds with boundary).

Let Q^n denote the standard unit n -cube in $\mathbb{R}^n$, i.e.,

$$Q^n = \{\mathbf{x} \in \mathbb{R}^n \mid 0 \leq x^i \leq 1 \text{ for all } i\}. \quad (105)$$

The boundary of Q^n consists of $2n$ sides, which we can treat as embeddings of the standard unit $(n-1)$ -cube Q^{n-1} . Hence ∂Q^n is a cubical chain, and the boundary of every cubical chain (made on standard unit cubes) is again a cubical chain of the same type. So it is a convenient class of chains, closed under the action of ∂ .

Let us describe the orientations induced on the sides of Q^n . (For all n we take the orientation of Q^n by the coordinates $x^1, \dots, x^n$.) Fix some $i = 1, \dots, n$. By setting x^i to 1 or to 0 we get two sides, which we denote $j_i^\pm$ where plus denotes $x^i = 1$ and minus denotes $x^i = 0$. We have the embeddings $j_i^\pm: Q^{n-1} \rightarrow Q^n$ where $x^1 = y^1, \dots, x^{i-1} = y^{i-1}, x^{i+1} = y^i, \dots, x^n = y^{n-1}$ and $x^i = 0$ or $x^i = 1$ (we denoted by y^i the coordinates on Q^{n-1}). The standard orientation of Q^{n-1} corresponds to the orientation by the basis $\mathbf{e}_1, \dots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \dots, \mathbf{e}_n$, i.e., the standard basis in $\mathbb{R}^n$ with the i -th vector omitted. Notice that $\mathbf{e}_i$ points outwards for j_i^+ and inwards for j_i^- . To get the induced orientation on $j_i^\pm$ we have to compare the orientation of $\pm \mathbf{e}_i, \mathbf{e}_1, \dots, \mathbf{e}_{i-1}, \mathbf{e}_{i+1}, \dots, \mathbf{e}_n$ with that of $\mathbf{e}_1, \dots, \mathbf{e}_n$. It follows that we have to take $(-1)^{i-1}$ for j_i^+ and $-(-1)^{i-1}$ for j_i^- . We have arrived at a statement:

Proposition 5.2. *The boundary of the cube Q^n is given by the formula:*

$$\partial Q^n = \sum_{i=1}^n (-1)^{i-1} (j_i^+ - j_i^-). \quad (106)$$

Problem 5.4. Check that $\partial(\partial Q^n) = 0$ as a cubical chain (made of maps $Q^{n-2} \rightarrow Q^n$).

Discussion: particular cases, recollection of the meaning of d and exterior product.

6 Classical integral theorems

Historical remark.

6.1 Forms corresponding to a vector field

In this section we consider $\mathbb{R}^n$ as a Euclidean space (see Remark 1.2). In particular, in the standard coordinates the scalar product of vectors $\mathbf{X}$ and $\mathbf{Y}$, which we denote $(\mathbf{X}, \mathbf{Y})$ or $\mathbf{X} \cdot \mathbf{Y}$, is given by the formula

$$(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \cdot \mathbf{Y} = X^1 Y^1 + \dots + X^n Y^n = \sum X^i Y^i. \quad (107)$$

We call coordinates in which the scalar product is given by (107) *Cartesian coordinates*. In particular, the standard coordinates are Cartesian. It is not difficult to see that all Cartesian coordinates are obtained from the standard coordinates by changes of variables of the form

$$y^i = \sum A_j^i x^j + b^i \quad (108)$$

where the matrix $A = (A_j^i)$ is orthogonal.

In an arbitrary coordinate system the expression for the scalar product will be more complicated.

Example 6.1. Consider polar coordinates in $\mathbb{R}^2$. To them correspond the basis $\mathbf{e}_r, \mathbf{e}_\varphi$:

$$\mathbf{e}_r = (\cos \varphi, \sin \varphi) \quad (109)$$

$$\mathbf{e}_\varphi = (-r \sin \varphi, r \cos \varphi). \quad (110)$$

For vectors $\mathbf{X} = X^1 \mathbf{e}_r + X^2 \mathbf{e}_\varphi$, $\mathbf{Y} = Y^1 \mathbf{e}_r + Y^2 \mathbf{e}_\varphi$ we obtain

$$\begin{aligned} (\mathbf{X}, \mathbf{Y}) &= (X^1 \mathbf{e}_r + X^2 \mathbf{e}_\varphi, Y^1 \mathbf{e}_r + Y^2 \mathbf{e}_\varphi) = \\ &= X^1 Y^1 (\mathbf{e}_r, \mathbf{e}_r) + X^1 Y^2 (\mathbf{e}_r, \mathbf{e}_\varphi) + X^2 Y^1 (\mathbf{e}_\varphi, \mathbf{e}_r) + X^2 Y^2 (\mathbf{e}_\varphi, \mathbf{e}_\varphi). \end{aligned}$$

Since from (109), (110) we have $(\mathbf{e}_r, \mathbf{e}_r) = 1$, $(\mathbf{e}_r, \mathbf{e}_\varphi) = 0$, and $(\mathbf{e}_\varphi, \mathbf{e}_\varphi) = r^2$, we finally obtain

$$(\mathbf{X}, \mathbf{Y}) = X^1 Y^1 + r^2 X^2 Y^2. \quad (111)$$

This is the formula for the scalar product in polar coordinates.

In general, if $\mathbf{X} = \sum X^i \mathbf{e}_i$ in the basis $\mathbf{e}_i = \frac{\partial \mathbf{x}}{\partial x^i}$ associated with a coordinate system $x^1, \dots, x^n$, and similarly for $\mathbf{Y}$, then

$$(\mathbf{X}, \mathbf{Y}) = \mathbf{X} \cdot \mathbf{Y} = \sum X^i Y^j g_{ij} \quad (112)$$

where $g_{ij} = (\mathbf{e}_i, \mathbf{e}_j)$. These coefficients in general depend on point, and the scalar product makes sense only for vectors attached to the same point.

Example 6.2. In polar coordinates the matrix (g_{ij}) is

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}. \quad (113)$$

Problem 6.1. For spherical coordinates r, θ, φ in $\mathbb{R}^3$ where

$$x = r \sin \theta \cos \varphi \quad (114)$$

$$y = r \sin \theta \sin \varphi \quad (115)$$

$$z = r \cos \theta \quad (116)$$

calculate pairwise scalar products of $\mathbf{e}_r, \mathbf{e}_\theta, \mathbf{e}_\varphi$ and show that

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}, \quad (117)$$

so

$$(\mathbf{X}, \mathbf{Y}) = X^1 Y^1 + r^2 X^2 Y^2 + r^2 \sin^2 \theta X^3 Y^3. \quad (118)$$

After these preliminaries we can pass to the main topic. Consider *vector fields* in $\mathbb{R}^n$, i.e., smooth functions associating with every point $\mathbf{x}$ in some open domain $U \subset \mathbb{R}^n$ a vector $\mathbf{X}(\mathbf{x})$ attached to this point: $\mathbf{x} \mapsto \mathbf{X}(\mathbf{x})$. We can visualize this as arrows attached to all points of U and smoothly varying with points. Treating them as velocity vectors, we obtain a picture of a vector field as a “flow” in U , so that each vector $\mathbf{X}(\mathbf{x})$ represents the velocity of the flow at the point $\mathbf{x}$. This hydrodynamical picture is very helpful.

With every vector field $\mathbf{X}$ in $\mathbb{R}^n$ are associated two differential forms:

(1) a 1-form denoted $\mathbf{X} \cdot d\mathbf{r}$, called the *circulation form* of $\mathbf{X}$,

(2) an $(n-1)$ -form denoted $\mathbf{X} \cdot d\mathbf{S}$, called the *flux form* of $\mathbf{X}$.

Before giving precise definitions, let us give a rough idea. Suppose we visualize $\mathbf{X}$ as a flow of some fluid. If γ is a path (or a 1-chain), it is natural to look for the measure of fluid that circulates along γ in a unit of time. Likewise, for an $(n-1)$ -dimensional surface in $\mathbb{R}^n$ (or an $(n-1)$ -chain) it is natural to look for the measure of fluid that passes across the surface in a unit of time. The answers to these questions are given by the integrals of the forms $\mathbf{X} \cdot d\mathbf{r}$ and $\mathbf{X} \cdot d\mathbf{S}$ respectively, hence the names. (“Flux” in Latin means “flow”.)

Definition 6.1. The *circulation form* corresponding to a vector field $\mathbf{X}$, notation: $\mathbf{X} \cdot d\mathbf{r}$ (or $\mathbf{X} \cdot d\mathbf{x}$), is a 1-form that on every vector $\mathbf{Y}$ takes the value $(\mathbf{X}, \mathbf{Y})$, the scalar product of $\mathbf{X}$ and $\mathbf{Y}$:

$$\langle \mathbf{X} \cdot d\mathbf{r}, \mathbf{Y} \rangle = (\mathbf{X}, \mathbf{Y}). \quad (119)$$

We immediately conclude that in Cartesian coordinates, if $\mathbf{X} = \sum X^i \mathbf{e}_i$, then

$$\mathbf{X} \cdot d\mathbf{r} = \sum X^i dx^i. \quad (120)$$

Indeed, the l.h.s. of (119) is always $\sum \omega_i Y^i$ in arbitrary coordinates (where $\mathbf{X} \cdot d\mathbf{r} = \omega = \sum \omega_i dx^i$). Comparing with (107), we get (120). In general coordinates we similarly obtain

$$\mathbf{X} \cdot d\mathbf{r} = \sum_{i,j} g_{ij} X^i dx^j \quad (121)$$

where $g_{ij} = (\mathbf{e}_i, \mathbf{e}_j)$.

Example 6.3. In the plane, if $\mathbf{X}$ is given in Cartesian coordinates x, y as $\mathbf{X} = X^1 \mathbf{e}_1 + X^2 \mathbf{e}_2$, then

$$\mathbf{X} \cdot d\mathbf{r} = X^1 dx + X^2 dy.$$

If $\mathbf{X}$ is given in polar coordinates r, φ as $\mathbf{X} = X^1 \mathbf{e}_r + X^2 \mathbf{e}_\varphi$ (the coefficients X^1, X^2 now have a different meaning), then

$$\mathbf{X} \cdot d\mathbf{r} = X^1 dr + r^2 X^2 d\varphi.$$

The correspondence between vectors fields and 1-forms given by $\mathbf{X} \mapsto \mathbf{X} \cdot d\mathbf{r}$ is invertible. Any 1-form $\omega = \sum \omega_i dx^i$ is the circulation form for a unique vector field $\mathbf{X}$. All we have to do is to solve the equation (121) for the coefficients X^i . We obtain

$$\mathbf{X} = \sum_{i,j} g^{ij} \omega_j \mathbf{e}_i \quad (122)$$

where g^{ij} (with upper indices) are the coefficients of the inverse matrix for (g_{ij}) . An important example is given by the notion of the gradient of a function.

Definition 6.2. The *gradient* of a function f , notation: $\text{grad } f$, is the vector field corresponding to the 1-form df :

$$\text{grad } f \cdot d\mathbf{r} = df. \quad (123)$$

Example 6.4. In Cartesian coordinates in $\mathbb{R}^n$

$$\text{grad } f = \frac{\partial f}{\partial x^1} \mathbf{e}_1 + \dots + \frac{\partial f}{\partial x^n} \mathbf{e}_n. \quad (124)$$

Example 6.5. In polar coordinates in $\mathbb{R}^2$

$$\text{grad } f = \frac{\partial f}{\partial r} \mathbf{e}_r + \frac{1}{r^2} \frac{\partial f}{\partial \varphi} \mathbf{e}_\varphi. \quad (125)$$

Example 6.6. In spherical coordinates in $\mathbb{R}^3$

$$\text{grad } f = \frac{\partial f}{\partial r} \mathbf{e}_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \mathbf{e}_\theta + \frac{1}{r^2 \sin^2 \theta} \frac{\partial f}{\partial \varphi} \mathbf{e}_\varphi. \quad (126)$$

We see that while df has a universal form in all coordinate systems, the expression for $\text{grad } f$ depends on particular coordinates.

Now we shall define the flux form $\mathbf{X} \cdot d\mathbf{S}$. To do so, we have to make a digression and discuss volumes in a Euclidean spaces. Let $x^1, \dots, x^n$ be Cartesian coordinates. The *volume* of any bounded domain $D \subset \mathbb{R}^n$ is defined as

$$\text{vol } D := \int_D dx^1 \dots dx^n. \quad (127)$$

It is clear that the definition does not depend on a choice of Cartesian coordinates if the orientation is not changed. (Indeed, the Jacobian is $\det A$ for A in (108), which is $+1$ or -1 , since A is orthogonal.) Hence we have actually defined the “oriented volume” of an oriented domain D , the absolute value of which is the usual volume. For $n = 2$ volume is called *area*. (For $n = 1$ “volume” is length.)

Example 6.7. Let $\Pi(\mathbf{a}, \mathbf{b})$ be a parallelogram spanned by vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^2$. Then by a direct calculation of the integral (check!)

$$\text{area } \Pi(\mathbf{a}, \mathbf{b}) = \begin{vmatrix} a^1 & a^2 \\ b^1 & b^2 \end{vmatrix} \quad (128)$$

if $\mathbf{a} = a^1 \mathbf{e}_1 + a^2 \mathbf{e}_2$, $\mathbf{b} = b^1 \mathbf{e}_1 + b^2 \mathbf{e}_2$ in Cartesian coordinates.

Obviously, this generalizes to an arbitrary n : in Cartesian coordinates the volume of parallelepiped in $\mathbb{R}^n$ spanned by vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ is

$$\text{vol } \Pi(\mathbf{a}_1, \dots, \mathbf{a}_n) = \begin{vmatrix} a_1^1 & \dots & a_1^n \\ \dots & \dots & \dots \\ a_n^1 & \dots & a_n^n \end{vmatrix}. \quad (129)$$

It is possible, starting from (128), (129) to give an “intrinsic” expression for this volume, entirely in terms of lengths of the vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ and angles between them, i.e., in terms of the pairwise scalar products of $\mathbf{a}_i$.

Lemma 6.1. For all n ,

$$(\text{vol } \Pi(\mathbf{a}_1, \dots, \mathbf{a}_n))^2 = \begin{vmatrix} \mathbf{a}_1 \cdot \mathbf{a}_1 & \dots & \mathbf{a}_1 \cdot \mathbf{a}_n \\ \dots & \dots & \dots \\ \mathbf{a}_n \cdot \mathbf{a}_1 & \dots & \mathbf{a}_n \cdot \mathbf{a}_n \end{vmatrix}. \quad (130)$$

At the r.h.s. stands the determinant made of all pairwise scalar products of $\mathbf{a}_i$.

This lemma can be proved by induction in n . It is clear for $n = 1$ ($|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a}$). For the inductive step one can notice that both sides does not change if to one of the vectors is added a linear combination of the others. We skip the remaining details.

Problem 6.2. Verify (130) directly for $n = 2$, by explicitly calculating the determinant in the r.h.s. in Cartesian coordinates.

The usefulness of Lemma 6.1 is double. First, it gives us a coordinate-free formula for the volume of a parallelepiped (and the area of a parallelogram). Second, because it does not involve coordinates, it is applicable to a system of k vectors in n -dimensional space for $k \leq n$ (as any such system is contained in a k -dimensional subspace). It gives in this case a formula for a k -volume in n -space. For example, it gives a formula for the area of an arbitrary parallelogram in $\mathbb{R}^3$.

Corollary 6.1. *In arbitrary coordinates the volume of a domain D is*

$$\text{vol } D := \int_D \sqrt{g} dx^1 \dots dx^n \quad (131)$$

where $g = \det(g_{ij})$, and $g_{ij} = \mathbf{e}_i \cdot \mathbf{e}_j$.

Indeed, when we calculate the integral in (131), the terms in the integral sum are the volumes of small parallelepipeds spanned by the vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ multiplied by $\Delta x^1 \dots \Delta x^n$ (small increments). We calculate each of the small volumes using formula (130) and add. Passing to the limit gives (131).

Hence the n -form

$$dV := \sqrt{g} dx^1 \dots dx^n \quad (132)$$

is the *volume form*. Its integral over a bounded domain gives the volume of the domain. In particular, in any Cartesian coordinates we have $g = 1$, and we return to the original definition of volume.

Similarly we can introduce a “ k -dimensional area” (or “ k -dimensional volume”) element dS for any k -dimensional surface in $\mathbb{R}^n$. Standard notation: dS (regardless of k). It is a k -form that “lives” on a surface, i.e., is a k -form written in terms of parameters on a surface. For any choice of parametrization, say, by variables $u^1, \dots, u^k$, the area element dS has the universal appearance

$$dS := \sqrt{h} du^1 \dots du^k \quad (133)$$

if we denote by h the determinant made of pairwise scalar products of the vectors $\mathbf{e}_i := \frac{\partial \mathbf{x}}{\partial u^i}$.

Example 6.8. For the sphere of radius R with center at O in $\mathbb{R}^3$

$$dS = R^2 \sin \theta d\theta d\varphi.$$

Remark 6.1. Volume element dV and area element dS are not differentials of any “ V ” or “ S ”, in spite of their notation. This notation is traditional and has a meaning of a “small element” giving vol or area after integration.

Now we can define the flux form $\mathbf{X} \cdot d\mathbf{S}$ for an arbitrary vector field $\mathbf{X}$ in $\mathbb{R}^n$.

Definition 6.3. The *flux form* corresponding to a vector field $\mathbf{X}$, notation: $\mathbf{X} \cdot d\mathbf{S}$, is an $(n-1)$ -form in Cartesian coordinates equal to

$$\mathbf{X} \cdot d\mathbf{S} = X^1 dx^2 dx^3 \dots dx^n - X^2 dx^1 dx^3 \dots dx^n + \dots + (-1)^{n-1} X^n dx^1 dx^2 \dots dx^{n-1}, \quad (134)$$

Proposition 6.1. The pull-back of the flux form $\mathbf{X} \cdot d\mathbf{S}$ for a vector field $\mathbf{X}$ in $\mathbb{R}^n$ by any $(n-1)$ -path $\Gamma: (u^1, \dots, u^{n-1}) \mapsto \mathbf{x}(u^1, \dots, u^{n-1})$ equals

$$\text{vol } \Pi(\mathbf{X}, \mathbf{e}_1, \dots, \mathbf{e}_{n-1}) du^1 \dots du^{n-1}.$$

where $\mathbf{e}_i = \frac{\partial \mathbf{x}}{\partial u^i}$, and $\text{vol } \Pi(\mathbf{a}, \dots, \mathbf{c})$ denotes the volume of a parallelogram spanned by vectors $\mathbf{a}, \dots, \mathbf{c}$.

Proof for $n = 3$. Let $\Gamma: \mathbf{x} = \mathbf{x}(u, v)$ be the 2-path in question. In Cartesian coordinates we have (since $\Gamma^*(dy dz) = (\partial_u y \partial_v z - \partial_v y \partial_u z) du dv$, etc.):

$$\begin{aligned} \Gamma^*(\mathbf{X} \cdot d\mathbf{S}) &= \Gamma^*(X^1 dy dz - X^2 dx dz + X^3 dx dy) = \\ &= X^1 \begin{vmatrix} \partial_u y & \partial_v y \\ \partial_u z & \partial_v z \end{vmatrix} du dv - X^2 \begin{vmatrix} \partial_u x & \partial_v x \\ \partial_u z & \partial_v z \end{vmatrix} du dv + X^3 \begin{vmatrix} \partial_u x & \partial_v x \\ \partial_u y & \partial_v y \end{vmatrix} du dv = \\ &= \begin{vmatrix} X^1 & X^2 & X^3 \\ \partial_u x & \partial_u y & \partial_u z \\ \partial_v x & \partial_v y & \partial_v z \end{vmatrix} du dv = \text{vol } \Pi(\mathbf{X}, \mathbf{e}_u, \mathbf{e}_v) du dv. \end{aligned}$$

□

(Similar proof works in any $\mathbb{R}^n$.)

.....

div

rotor

calculation in arb coord

ex in polar coordinates

6.2 The Ostrogradski–Gauss and classical Stokes theorems

The specialization of the general Stokes theorem for the flux form $\mathbf{X} \cdot d\mathbf{S}$ and the circulation form $\mathbf{X} \cdot d\mathbf{r}$ gives two classical integral theorems traditionally associated with the names of Ostrogradski, Gauss, and Stokes.

Let us fix some orientation in $\mathbb{R}^n$. This makes it possible to consider integrals of n -forms over any bounded domains $D \subset \mathbb{R}^n$ without ambiguity in sign.

Definition 6.4. Let S be an oriented surface of dimension $n - 1$ in $\mathbb{R}^n$ or an $(n - 1)$ -chain. The *flux* of a vector field $\mathbf{X}$ through S is defined as the integral of the flux form $\mathbf{X} \cdot d\mathbf{S}$ over S :

$$\int_S \mathbf{X} \cdot d\mathbf{S}.$$

The general Stokes theorem and the definitions of $\mathbf{X} \cdot d\mathbf{S}$ and $\operatorname{div} \mathbf{X}$ immediately imply

Theorem 6.1 (Ostrogradski–Gauss theorem). *The flux of a vector field $\mathbf{X}$ through the boundary of any bounded domain $D \subset \mathbb{R}^n$ equals the volume integral of the divergence of $\mathbf{X}$:*

$$\oint_{\partial D} \mathbf{X} \cdot d\mathbf{S} = \int_D \operatorname{div} \mathbf{X} \, dV. \quad (135)$$

Written in Cartesian coordinates, equation (135) up to a change in notation coincides with the version (104) of the general Stokes theorem (which makes sense without any Euclidean structure). Peculiar for a Euclidean space is the possibility to formulate it in terms of a vector field.

There is an extra statement giving an interpretation of the l.h.s. of (135) that helps to better understand its geometrical meaning. We need the notion of a unit normal for this.

Notice that for any $(n - 1)$ -dimensional surface S in $\mathbb{R}^n$ at every point of S we can consider a unit normal (a normal vector of unit length) $\mathbf{n}$. It is defined up to a sign: $\pm \mathbf{n}$. However, if the surface is oriented, the direction of $\mathbf{n}$ is defined uniquely. Recall that we have fixed an orientation in the ambient space. The condition is that the basis $\mathbf{n}, \mathbf{g}_1, \dots, \mathbf{g}_{n-1}$ gives the orientation of $\mathbb{R}^n$ if the basis $\mathbf{g}_1, \dots, \mathbf{g}_{n-1}$ gives the orientation of the surface S . Conversely, if $\mathbf{n}$ is given, this fixes an orientation of S , from the chosen orientation of $\mathbb{R}^n$.

Example 6.9. If $S = \partial D$ with the induced orientation (see Definition 5.1), then $\mathbf{n}$ must point outwards. Hence $\mathbf{n}$ is the *outward normal*, and Definition 5.1 is often referred to as the “outward normal rule”.

Example 6.10. In $\mathbb{R}^3$, if a piece of a surface is given in the parametric form as $\mathbf{x} = \mathbf{x}(u, v)$, then the parameters u, v define an orientation of the surface via the basis $\mathbf{e}_u = \frac{\partial \mathbf{x}}{\partial u}, \mathbf{e}_v = \frac{\partial \mathbf{x}}{\partial v}$ of the tangent plane. The unit normal corresponding to this orientation is given by

$$\mathbf{n} = \frac{\mathbf{e}_u \times \mathbf{e}_v}{|\mathbf{e}_u \times \mathbf{e}_v|}. \quad (136)$$

Indeed, the rule defining the cross product is that $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ should give a positive basis, and this is equivalent to $\mathbf{a} \times \mathbf{b}, \mathbf{a}, \mathbf{b}$ giving a positive basis.

(Actually, this is valid for arbitrary dimension n if the notion of a “cross product” is suitably generalized, so that it takes as arguments $n - 1$ vectors instead of 2 vectors in $\mathbb{R}^3$.)

Proposition 6.2. *The restriction of the flux form $\mathbf{X} \cdot d\mathbf{S}$ to any oriented surface of dimension $n - 1$ in $\mathbb{R}^n$ equals $(\mathbf{X} \cdot \mathbf{n}) dS$ where $\mathbf{n}$ is the corresponding unit normal and dS is the area element.*

Proof (for $n = 3$). Immediately follows from Proposition 6.1 and the formula for the normal vector. $\square$

It follows that for the same magnitude of $\mathbf{X}$, the “elementary” flux of $\mathbf{X}$ through a surface S near some point is maximal for $\mathbf{X}$ normal to S , and equals zero for $\mathbf{X}$ tangent to S , which is exactly our intuitive picture of a “flow across the surface S ”.

Example 6.11. Find the flux of a “constant flow” along the x -axis,

$$\mathbf{X} = a \mathbf{e}_1$$

across a unit square in the plane P_α passing through the y -axis with an orientation specified by a unit normal $\mathbf{n} = (\cos \alpha, 0, \sin \alpha)$ (so the plane is at angle α with the z -axis). By Proposition 6.2, $\mathbf{X} \cdot d\mathbf{S} = (a \cos \alpha) dS$; thus the flux is $a \cos \alpha$. It takes the maximal value a when $\alpha = 0$, and when we rotate the plane the flux decreases to 0 for $\alpha = \pi/2$, becomes negative, and takes the value $-a$ for $\alpha = \pi$, when the orientation is “opposite to the flow”.

Example 6.12. Consider the flux of the vector field

$$\mathbf{E} = -\frac{\mathbf{r}}{r^3} \quad (137)$$

in $\mathbb{R}^3$ (the “Coulomb force”) through the sphere of radius R oriented by the outward normal. The Ostrogradski–Gauss theorem is not applicable because $\mathbf{E}$ is not defined at the origin O . (There is a trick overcoming this, see Example 5.16.) Using Proposition 6.2 we can evaluate the flux easily. Indeed, as $\mathbf{r}$ points in the direction of the outward normal, we have $(\mathbf{E} \cdot \mathbf{n}) dS = -RR^{-3} dS = -R^{-2} dS$ (as $r = R$ on the sphere). Hence

$$\oint_{S_R} \mathbf{E} \cdot d\mathbf{S} = \oint_{S_R} (\mathbf{E} \cdot \mathbf{n}) dS = -R^{-2} \oint_{S_R} dS = -R^{-2} \text{area } S_R = -4\pi. \quad (138)$$

We see that remarkably the flux does not depend on radius. The explanation is that the form $-r^{-3} \mathbf{r} \cdot d\mathbf{S}$ is closed, or, equivalently, that $\text{div}(-r^{-3} \mathbf{r}) = 0$, for $\mathbf{r} \neq 0$, in $\mathbb{R}^3$.

From the Ostrogradski–Gauss theorem follows the “integral definition” of the divergence:

$$\operatorname{div} \mathbf{X}(\mathbf{x}_0) = \lim_{D \rightarrow \mathbf{x}_0} \frac{\oint_{\partial D} \mathbf{X} \cdot d\mathbf{S}}{\operatorname{vol} D}. \quad (139)$$

Here $\mathbf{x}_0 \in D$ and $D \rightarrow \mathbf{x}_0$ means that the domain D “shrinks” to a point $\mathbf{x}_0$. Thus the divergence at $\mathbf{x}_0$ measures the intensity of a “source” of the flow at the point $\mathbf{x}_0$. If it is negative, the “source” is actually a “sink”. All these concepts come from the hydrodynamical interpretation.

Another statement following from the general Stokes theorem and which gave to it the name is the “classical Stokes theorem”.

$$\text{In Cartesian coordinates } \operatorname{curl} \mathbf{X} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \partial_1 & \partial_2 & \partial_3 \\ X^1 & X^2 & X^3 \end{vmatrix} \quad (\text{this works only in } \mathbb{R}^3).$$

Theorem 6.2 (Classical Stokes theorem). *The circulation of a vector field over the boundary of any oriented surface or 2-chain C in $\mathbb{R}^3$ equals the flux of the curl of $\mathbf{X}$ through C :*

$$\oint_{\partial C} \mathbf{X} \cdot d\mathbf{r} = \int_C \operatorname{curl} \mathbf{X} \cdot d\mathbf{S}.$$